

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

THIS PAGE BLANK (USPTO)



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 108576

TO: Bao-Qun Li
Location: cm1/7e07/8e12
Art Unit : 1648
Friday, November 28, 2003

Case Serial Number: 09/973025

From : Susan Hanley
Location: Biotech-Chem Library
CM1 6B05
Phone: 305-4053

susan.hanley@uspto.gov

Search Notes

THIS PAGE BLANK (USPTO)

108576

STIC-Biotech/ChemLib

From: Li, Bao-Qun
Sent: Tuesday, November 18, 2003 3:11 PM
To: STIC-Biotech/ChemLib

RECEIVED

NOV 19 2003

Please do the oligmer peptide homology and interference search for the SEq ID NO: 72-90 and 91 of Application SN. 09/973,025. Thanks. CM1, 8E12.

(STIC)

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:22 ; Search time 32.1 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-72

Perfect score: 20
Sequence: 257
257
416

Scoring table:
OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_19Jun03;*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	16	AA1984432
2	20	100.0	20	17	AA1984432
3	20	100.0	20	23	AA018701
4	20	100.0	23	20	AA199518
5	20	100.0	26	14	AA1941142
6	20	100.0	36	14	AA1941140
7	20	100.0	144	14	AA1933998
8	20	100.0	174	14	AA1934441
9	20	100.0	319	15	AA1945330

10	20	100.0	513	13	AA1924086
11	20	100.0	527	13	AA1925136
12	20	100.0	733	14	AA1938278
13	20	100.0	2894	13	AA1924440
14	20	100.0	2894	16	AA1970230
15	20	100.0	3011	15	AA1966935
16	19	95.0	35	13	AA1920759
17	13	65.0	30	23	AA194624
18	13	65.0	30	23	AA194625
19	13	65.0	91	23	AA197256
20	13	65.0	350	23	AA198931
21	13	65.0	363	24	AA192880
22	13	65.0	363	24	AA195567
23	13	65.0	402	14	AA1934439
24	13	65.0	409	14	AA1933995
25	13	65.0	1006	18	AA192715
26	13	65.0	1997	23	AA194802
27	13	65.0	2010	23	AA194801
28	13	65.0	3011	13	AA1922154
29	13	65.0	3011	19	AA197397
30	13	65.0	3011	19	AA197398
31	13	65.0	3011	20	AA198020
32	13	65.0	3011	22	AA191169
33	13	65.0	3011	22	AA1959173
34	13	65.0	3011	23	AA1999290
35	13	65.0	3011	23	AA1979221
36	13	65.0	3011	23	AA191988
37	13	65.0	3011	23	AA194597
38	13	65.0	3011	24	AA1971460
39	13	65.0	3011	24	AA1926784
40	13	65.0	3012	23	AA1999289
41	13	65.0	3180	24	AA1973195
42	13	65.0	3835	22	AA191167
43	13	65.0	5985	23	AA194799
44	12	60.0	31	21	AA1977233
45	12	60.0	31	21	AA1977237

ALIGNMENTS

RESULT 1
ID AA1984432 standard; peptide; 20 AA.
AC AA1984432;
DT 06-JAN-1997 (first entry)
XX Hepatitis C virus E2 region (397-416) peptide NS1-7*.
DE Hepatitis C virus; HCV; immunogen; E2 region; immunodominant;
KW T cell epitope; vaccine.
XX Hepatitis C virus.
OS WO19512677-A2.
PN 11-MAY-1995.
PD 28-OCT-1994; 94WO-EP03555.
PP 04-NOV-1993; 93EP-0402718.
PR (INNO-) INNOGENETICS NV.
XX Deleys R, Leroux-Roels G, Maertens G;
XX WPI; 1995-193822/25.
XX Hepatitis C virus immunogenic polypeptide contg. a T-cell
stimulating epitope - from core, E1, E2 and NS3 regions, useful in
production of vaccines, therapeutic agents, etc.

XX PS Claim 26; Page 71; 105pp; English.

CC Polypeptides comprising 8-32 amino acids from the HCV E2 region

CC sequence in AAR84356 (spanning positions 397-428) and containing a

CC T-cell stimulating epitope are used in HCV immunogenic compositions.

CC The present sequence is a specifically claimed example of such a

CC T-cell epitope-containing peptide.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 16; Length 20;

Best Local Similarity 100.0%; Pred. No. 3.5e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLINT 20

Db 1 SGLVSLFTPGAKQNIQLINT 20

RESULT 2

AAR90995

ID AAR90995 standard; peptide; 20 AA.

XX AC AAR90995;

XX DT 25-SEP-1996 (first entry)

XX DE HCV E2 peptide E2-67 for competition studies.

XX KW HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;

XX KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.

XX OS Synthetic.

XX PN WO9604385-A2.

XX PD 15-FEB-1996.

XX PF 31-JUL-1995; 95WO-EP03031.

XX PR 29-JUL-1994; 94EP-0870132.

XX PA (INNO-) INNOGENETICS NV.

XX PI Bosman F, Buyse M, De Martynoff G, Maertens G;

XX WPI; 1996-129401/13.

XX PT Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope

XX PT proteins - in presence of disulphide bond cleavage agent, to

XX PT produce proteins suitable for direct use in vaccines or diagnostic

XX PT assays of HCV

XX PS Claim 29; Page 67; 146pp; English.

XX CC AAR90965-R90980 and AAR90995-R91015 represent synthetic hepatitis C

XX CC virus (HCV) E1 and E2 peptides used in competition studies. This

XX CC sequence represents a synthetic E2 peptide, and corresponds to residues

XX CC 397-416 of the E2 protein sequence. These sequences are useful for in

XX CC vitro monitoring of HCV disease, or prognosis of the response to

XX CC interferon treatment of patients suffering from HCV infection. These

XX CC sequences compete with the proteins produced by AAT12704-T12709 and

XX CC AAT12961-T12974, which are included in vectors for the production of

XX CC recombinant E1, E2, and E1/E2. The recombinant proteins can then be

XX CC isolated and purified by carrying out a disulphide bond cleavage, or a

XX CC reduction step with a disulphide bond cleavage agent, after lysis of

XX CC recombinant host cells. The constructs containing the purified HCV

XX CC envelope proteins can be used for vaccinating humans against HCV, for in

XX CC vitro detection of HCV antibodies in a sample, and in a serotyping assay

XX CC for detecting one or more serological types of HCV present in a

XX CC biological sample. The constructs can also be immobilised on a solid

XX CC substrate and incorporated into a reversed phase hybridisation assay for

CC determining the presence or the genotype of HCV. The new purification

CC method preserves the conformation of the recombinantly expressed E1, E2

CC and E1/E2, and eliminates contaminating proteins. Antigens isolated

CC using this method are more reactive with human sera than those isolated

CC by known techniques.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 17; Length 20;

Best Local Similarity 100.0%; Pred. No. 3.5e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLINT 20

Db 1 SGLVSLFTPGAKQNIQLINT 20

RESULT 3

AAO18701

ID AAO18701 standard; Peptide; 20 AA.

XX AC AAO18701;

XX DT 24-OCT-2002 (first entry)

XX DE Hepatitis C virus E2 protein derived peptide E2-67.

XX KW Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide;

XX KW immunostimulant; vaccine.

XX OS Hepatitis C virus.

XX PN WO200255548-A2.

XX PD 18-JUL-2002.

XX PF 11-JAN-2002; 2002WO-EP00219.

XX PR 11-JAN-2001; 2001US-260699P.

XX PR 30-AUG-2001; 2001US-315768P.

XX PA (INNO-) INNOGENETICS NV.

XX PI Maertens G, Bosman F, Buyse M;

XX WPI; 2002-599657/64.

XX PT New therapeutic vaccine compositions comprising at least one purified

XX PT recombinant hepatitis C virus (HCV) single or specific oligomeric

XX PT recombinant envelope protein E1 or E2, useful for immunizing humans

XX PT from HCV infection

XX PS Claim 4; Page 225; 243pp; English.

XX CC The present invention relates to new therapeutic vaccine compositions for

XX CC inducing hepatitis C virus (HCV)-specific antibodies, comprising a

XX CC composition containing at least one purified recombinant HCV single or

XX CC specific oligomeric recombinant envelope proteins selected from an E1 and

XX CC an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are

XX CC useful for inducing HCV-specific antibodies or for immunising humans

XX CC against HCV. The recombinant HCV E1 and/or E2 proteins are useful as

XX CC vaccines or therapeutics, in HCV screening and confirmatory antibody

XX CC tests, for raising antibodies, in the preparation of medicament, and for

XX CC in vitro monitoring of HCV disease or prognosing the response to

XX CC treatment of patients suffering from HCV infection. The present sequence

XX CC is a peptide derived from the proteins of the invention.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 23; Length 20;

Best Local Similarity 100.0%; Pred. No. 3.5e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLINT 20
 Db 1 SGLVSLFTPGAKQNIQLINT 20

RESULT 4

AAAY39518
 ID AAY39518 standard; peptide; 23 AA.

XX AC AAY39518;

XX DT 22-NOV-1999 (first entry)

XX HCV E2 protein residues 394-416.

XX HCV; E1 protein; E2 protein; epitope; hepatitis c virus; antibody;
 KW detection; HCV protein antigen.

XX Hepatitis c virus.

XX EP947525-A1.

XX 06-OCT-1999.

XX 27-MAR-1998; 98EP-0870060.

XX 27-MAR-1998; 98EP-0870060.

XX (INNO-) INNOGENETICS NV.

XX WPI; 1999-542955/46.

XX New anti-Hepatitis C virus (HCV) antibodies useful for in situ
 PT detection of HCV -

XX Example 4; Page 9; 32pp; English.

XX This sequence represents residues 394-416 of the hepatitis c virus (HCV)
 CC E2 protein. The invention relates to an antibody (Ab) or its derivative
 CC that specifically binds to the C-terminal region of the HCV E1 protein
 CC (amino acids (aa) 227-383) or the N-terminal region of the HCV E2 protein
 CC (aa 384-450). The Ab allows the in situ detection of HCV protein
 CC antigens. The Ab is used to detect the HCV E1 or E2 proteins in human
 CC peripheral blood cells or liver tissue. It may also be used to detect HCV
 CC in other tissue and fluid samples such as serum, plasma, saliva, mucus
 CC and sections or biopsies such as from skin.

XX Sequence 23 AA;

Query Match 100.0%; Score 20; DB 20; Length 23;
 Best Local Similarity 100.0%; Pred. No. 3.8e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLINT 20
 Db 4 SGLVSLFTPGAKQNIQLINT 23

RESULT 5

AAAR41142
 ID AAR41142 standard; peptide; 26 AA.

XX AC AAR41142;

XX 25-MAR-2003 (updated)

XX DT 22-MAR-1994 (first entry)

XX HCV peptide XXG-2 (aa 393-416; E2/NS1 N-terminal).

XX Human immunodeficiency virus; HIV; hepatitis C virus; HCV;
 KW non-A non-B hepatitis; NANBH; human T-cell lymphotropic virus; HTLV;
 KW epitope; antibody; biotin; diagnosis; detection; vaccine.

Synthetic.

Key Location/Qualifiers

Modified-site 1

/note= "the N-terminal comprises (A)-(B)-(X)-Y; where

B= biotin;

X= biotinylation cpd. incorporated

Y= bond or linking gp(s). which

minimises steric hindrance.

where Y is not a bond it is pref. 1-10

residues of (same or different) glycine,

beta-alanine, 4-aminobutyric acid,

5-aminovaleic acid or 6-aminohexanoic acid;

parentesis around B and X indicate opt. presence

at the specified positions but B or X must be

present in at least one of the positions shown,

B interacts with the peptide to give a cpd.

with greater diagnostic sensitivity;

A (optional)= one or more amino acids, NH2 or

gp. which modifies the N-terminus;

Z= one or more amino acids, OH, NH2, or a

linkage involving either of these 2 gps."

Modified-site 26

/note= "the C-terminal comprises Y-(X)-Z"

WO9318054-A2.

16-SEP-1993.

08-MAR-1993; 93WO-EP00517.

06-MAR-1992; 92EP-0400598.

(INNO-) INNOGENETICS NV SA.

De LEYS, R.

WPI; 1993-303397/38.

New biotinylated peptide(s) corresp. to immuno-dominant
 epitope(s) - with increased antigenicity, useful in antibodies
 detection and vaccines against hepatitis C, HIV and HTLV

Claim 4; Page 90-98; 133pp; English.

Peptide compns. comprise at least one and pref. a combination of
 two, three, four or more biotinylated peptides chosen from the
 sequences given in AAR41058-R41166. The peptides represent
 immunologically important regions of viral proteins and are
 prepd. by solid phase peptide synthesis. The compns. are
 useful for the detection of antibodies to HCV, and/or HIV,
 and/or HTLV-I or II.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 26 AA;

Query Match 100.0%; Score 20; DB 14; Length 26;

Best Local Similarity 100.0%; Pred. No. 4.2e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLINT 20

Db 6 SGLVSLFTPGAKQNIQLINT 25

RESULT 6

AAAR41140

ID AAR41140 standard; peptide; 36 AA.

XX AC AAR41140;

XX DT 25-MAR-2003 (updated)

22-MAR-1994 (first entry)
 HCV peptide XXg (aa 383-416; E2/NS1 N-terminal).
 Human immunodeficiency virus; HIV; hepatitis C virus; HCV;
 non-A non-B hepatitis; NANBH; human T-cell lymphotropic virus; HTLV;
 epitope; antibody; biotin; diagnosis; detection; vaccine.
 Synthetic.

Key Location/Qualifiers
 Modified-site 1
 /note= "the N-terminal comprises (A) - (B) - (X) - Y; where
 B= biotin;
 X= biotinylation cpd. incorporated
 Y= bond or linking gp (s). which
 minimises steric hindrance,
 where Y is not a bond it is pref. 1-10
 residues of (same or different) glycine,
 beta-alanine, 4-aminobutyric acid,
 5-aminovaleric acid or 6-aminohexanoic acid;
 parentheses around B and X indicate opt. presence
 at the specified positions but B or X must be
 present in at least one of the positions shown,
 B interacts with the peptide to give a cpd.
 with greater diagnostic sensitivity;
 A (optional)= one or more amino acids, NH2 or
 Z= one or more amino acids, OH, NH2, or a
 linkage involving either of these 2 gps."

Modified-site 36
 /note= "the C-terminal comprises Y-(X)-Z"

WO9318054-A2.
 16-SEP-1993.
 08-MAR-1993; 93WO-EP00517.
 06-MAR-1992; 92EP-0400598.
 (INNO-) INNOGENETICS NV SA.
 De LEYS R;
 WPI; 1993-303397/38.
 New biotinylated peptide(s) corresp. to immuno-dominant
 epitope(s) - with increased antigenicity, useful in antibodies
 detection and vaccines against hepatitis C, HIV and HTLV
 Claim 4; Page 90-98; 133pp; English.
 Peptide compns. comprise at least one and pref. a combination of
 two, three, four or more biotinylated peptides chosen from the
 sequences given in AAR41058-R41166. The peptides represent
 immunologically important regions of viral proteins and are
 prepd. by solid phase peptide synthesis. The compns. are
 useful for the detection of antibodies to HCV, and/or HIV,
 and/or HTLV-I or II.
 (Updated on 25-MAR-2003 to correct PN field.)

Query Match 100.0%; Score 20; DB 14; Length 36;
 Best Local Similarity 100.0%; Pred. No. 5.4e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SGLVSLFTPGAKONIQLINT 20
 |||||
 DB 16 SGLVSLFTPGAKONIQLINT 35

RESULT 7

AAR33998
 ID AAR33998 standard; Protein; 144 AA.

XX
 AC AAR33998;

XX
 DT 25-MAR-2003 (updated)
 DT 26-JUL-1993 (first entry)

XX
 DE HC-J1 E2/NS1 protein.

XX
 KW Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
 KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
 KW domain; immunological; cross-reactive; envelope protein; vaccine;
 KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.

OS Synthetic.

XX
 PN WO9306126-A1.

XX
 PD 01-APR-1993.

XX
 PF 11-SEP-1992; 92WO-US07683.

XX
 PR 13-SEP-1991; 91US-0759575.

XX
 PA (CHIR) CHIRON CORP.

XX
 PI Houghton M, Weiner AJ;

XX
 DR WPI; 1993-117468/14.

XX
 PT Immuno-reactive hepatitis C virus polypeptide compns. - contg.
 PT at least 2 sequences from the first variable domain of distinct
 PT HCV isolates

PS Disclosure; Fig 3; 106pp; English.

XX
 CC The sequences given in AAR33992-002 represent a portion of the E2/NS1
 CC protein encoded by group I and group II HCV isolates, from amino acid
 CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
 CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
 CC 30 amino acids which shows large variation between nearly all isolates.
 CC This is an important immunoreactive domain. This putative envelope
 CC glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera
 CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
 CC flaviviruses, both of which confer protective immunity in hosts
 CC vaccinated with these polypeptides. It has been discovered that a
 CC number of important HCV epitopes vary among viral isolates and that
 CC these epitopes can be mapped to specific domains. This meant that
 CC immunologically cross-reactive polypeptides which focus on variable
 CC rather than constant domains can be produced. See also AAQ39134-48
 CC and AAR33982-91.
 CC (Updated on 25-MAR-2003 to correct PN field.)

SQ Sequence 144 AA;

Query Match 100.0%; Score 20; DB 14; Length 144;

Best Local Similarity 100.0%; Pred. No. 1.5e-12;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SGLVSLFTPGAKONIQLINT 20

|||||
 DB 28 SGLVSLFTPGAKONIQLINT 47

RESULT 8

AAR34441
 ID AAR34441 standard; Protein; 174 AA.

XX
 AC AAR34441;

XX

DT 25-MAR-2003 (updated)
 DT 09-AUG-1993 (first entry)
 XX
 DE Sequence of glycoprotein E2/NS1 in clone J1(JM).
 XX
 KW Hepatitis C virus; envelope protein; glycoprotein; E2/NS1;
 KW diagnostic reagent.
 XX
 OS Hepatitis C virus.
 XX
 PN EP537626-A1.
 PN
 PD 21-APR-1993.
 XX
 PF 08-OCT-1992; 92EP-0117191.
 XX
 PR 08-OCT-1991; 91JP-0260824.
 XX
 PA (NAHE-) NAT INST OF HEALTH.
 XX
 XX Harada S, Honda Y, Miyamura T, Saito I;
 XX WPI; 1993-127516/16.
 DR N-PSDB; AAQ40333.
 DR
 XX Diagnostic reagent for hepatitis C virus - comprises second
 PT envelope protein or first non-structural protein encoded by HCV
 PT gene and has sugar chain
 XX
 PS Claim 2; Pages 40-41; 58pp; English.
 XX
 CC Glycoprotein E2/NS1 is derived from the second envelope protein or
 CC first non-structural protein encoded by the genome of HCV. The
 CC nucleic acid is extracted from the serum of the patient of hepatitis
 CC C. The serum is pref. mixed with transfer RNA (tRNA) as a carrier
 CC of virus RNA. As a technique of cloning cDNA from the nucleic acid,
 CC it is preferred to use polymerase chain reaction method. In the
 CC reaction, any commercially available random primers or synthesized
 CC DNA having a base sequence similar to that of primer AS1 may be used
 CC as a primer. Representative examples of sense primers include S1.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 174 AA;
 Query Match 100.0%; Score 20; DB 14; Length 174;
 Best Local Similarity 100.0%; Pred. No. 1.8e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 SGLVSLFTPGAKQNIQLINT 20
 Db 58 SGLVSLFTPGAKQNIQLINT 77
 RESULT 9
 AAR45330
 ID AAR45330 standard; protein; 319 AA.
 XX
 AC AAR45330;
 XX
 XX 28-JUN-1994 (first entry)
 DT
 XX Anti-HCV antibody reactive protein #1.
 DE
 XX Hepatitis C virus; HCV; envelope; region; anti-HCV; antibody; vaccine.
 KW
 XX Synthetic.
 OS
 XX Key Location/Qualifiers
 FH 39..69
 FT Peptide /label= EPI
 FT /note= "Claim 1"
 FT
 XX JP05310786-A.
 PN

XX 22-NOV-1993.
 PD
 XX 30-APR-1992; 92JP-0111846.
 XX
 PF 30-APR-1992; 92JP-0111846.
 PR
 XX (OLYU) OLYMPUS OPTICAL CO LTD.
 PA
 XX WPI; 1994-002315/01.
 DR
 XX Peptide of HCV envelope region - having high reactivity against
 PT anti-hepatitis C virus antibody and having at least 6 amino acids
 PT
 XX Disclosure; Page 6-7; 11pp; Japanese.
 PS
 XX The sequences given in AAR45330-31 represent fragments derived from
 CC the hepatitis C virus (HCV) envelope region that are highly reactive
 CC against anti-HCV antibodies. These protein fragments may be used in
 CC vaccines against HCV. They have a high reactivity against anti-HCV
 CC antibodies regardless of the original HCV. They can be used to
 CC detect HCV infection but not the onset of infection.
 XX
 SQ Sequence 319 AA;
 Query Match 100.0%; Score 20; DB 15; Length 319;
 Best Local Similarity 100.0%; Pred. No. 2.8e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 SGLVSLFTPGAKQNIQLINT 20
 Db 206 SGLVSLFTPGAKQNIQLINT 225
 RESULT 10
 AAR24086
 ID AAR24086 standard; Protein; 513 AA.
 XX
 AC AAR24086;
 XX
 XX 09-NOV-1992 (first entry)
 DT
 XX NANB hepatitis virus strain HC-J1 gene product.
 DE
 XX non-A, non-B hepatitis virus; NANBH; HC-J5; PCR; amplification
 KW polymerase chain reaction; vaccine; antibody.
 XX
 XX Non-A, non-B hepatitis virus.
 OS
 XX EP485209-A.
 PN
 PD 13-MAY-1992.
 XX
 PF 07-NOV-1991; 91EP-0310297.
 XX
 PR 08-NOV-1990; 90JP-0304405.
 XX
 XX (IMMO) IMMUNO JAPAN INC.
 PA
 XX Nakamura T, Okamoto H;
 PI
 XX WPI; 1992-160959/20.
 DR
 XX N-PSDB; AAQ24466.
 DR
 XX Recombinant cDNA of NANBH virus strain HC-J5 and corresp.
 PT peptides - useful for diagnosis and in vaccines and immunological
 PT pharmaceuticals
 XX
 XX Disclosure; Page 19; 42pp; English.
 PS
 XX This sequence is the gene product from the genome of the non-A, non-B
 CC hepatitis virus (NANBH) strain HC-J1. This sequence was derived
 CC from the DNA sequence which was amplified by polymerase chain reaction.

CC The nucleotide sequences derived from this amplification can be used to
 CC detect NANBH virus infection which could not be detected by conventional
 CC methods. The detection kits allow highly specific and sensitive
 CC detection at an early phase of infection. This polypeptide sequence
 CC can be used for the manufacture of vaccines and immunological
 CC pharmaceuticals and also to produce antibodies specific to NANBH.
 XX
 XX

SQ Sequence 513 AA;

Query Match 100.0%; Score 20; DB 13; Length 513;
 Best Local Similarity 100.0%; Pred. No. 4.1e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGLVSLFTPGAKQNIQLINT 20
 |||||
 DB 397 SGLVSLFTPGAKQNIQLINT 416

RESULT 11
 AAR25136
 ID AAR25136 standard; Protein; 527 AA.

XX AC AAR25136;

DT 23-DEC-1992 (first entry)

DE HCV polypeptide 2.

KW Hepatitis C virus; blood transfusion.

XX Synthetic.

PN JP04159298-A.

XX 02-JUN-1992.

PF 19-OCT-1990; 90JP-0282431.

PR 19-OCT-1990; 90JP-0282431.

PA (OLYU) OLYMPUS OPTICAL CO LTD.

DR WPI; 1992-231947/28.

PT New peptides acting as antigenic analogues of human hepatitis C
 PT virus - useful for detecting HCV antibody positive patients

PS Disclosure; Fig 2; 14pp; Japanese.

XX The sequences given in AAR25135-36 are peptides from the hepatitis C
 CC virus (HCV) which are recognised by the peptide sequences given in
 CC AAR25130-24. These antigenic peptides can be used on their own or as
 CC a mixture two different peptides. Using these peptides, HCV antibody
 CC positive patients can be detected and hepatitis caused by blood
 CC transfusion can be prevented.

SQ Sequence 527 AA;

Query Match 100.0%; Score 20; DB 13; Length 527;
 Best Local Similarity 100.0%; Pred. No. 4.1e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGLVSLFTPGAKQNIQLINT 20
 |||||
 DB 411 SGLVSLFTPGAKQNIQLINT 430

RESULT 12
 AAR38278
 ID AAR38278 standard; Protein; 733 AA.

XX AAR38278;

XX

DT 21-OCT-1993 (first entry)
 XX
 DE NANB hepatitis virus HC-OM gene polypeptide P-733-1.
 XX
 KW Non-A, non-B; virus; polymerase chain reaction; detection;
 KW sensitive; specific; HCV; NANBH.
 XX
 OS Non-A, non-B hepatitis virus.
 XX JP05091884-A.
 PN 16-APR-1993.
 PD 10-APR-1991; 91JP-0196175.
 PF 12-JUN-1990; 90JP-0153401.
 PR 08-NOV-1990; 90JP-0304405.
 XX (NAKA/) NAKAMURA T.
 PA WPI; 1993-199637/25.
 DR N-PSDB; AAQ43888.
 XX
 PT Antigen related to non-A and non-B hepatitis virus - comprises
 PT non-translation region comprising 340 - 341 mols. of nucleotides,
 PT non-translation region comprising 1885 - 2551 mols. of
 PT nucleotides including region 1,149 and, etc.
 XX
 PS Claim 14; Page 32-35; 73pp; Japanese.

XX The sequence is that of NANB hepatitis virus HC-OM gene polypeptide
 CC P-733-1. It may be used in a system for detecting NANB hepatitis.
 CC This method is highly specific and sensitive, and can detect NANB
 CC hepatitis virus which could not be detected by conventional methods.
 XX

SQ Sequence 733 AA;

Query Match 100.0%; Score 20; DB 14; Length 733;
 Best Local Similarity 100.0%; Pred. No. 5.3e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGLVSLFTPGAKQNIQLINT 20
 |||||
 DB 397 SGLVSLFTPGAKQNIQLINT 416

RESULT 13

AAR24440
 ID AAR24440 standard; protein; 2894 AA.

XX AC AAR24440;

DT 25-MAR-2003 (updated)

DT 02-DEC-1992 (first entry)

XX Composite HCV HC-J1/CDC/CHI protein.

XX Hepatitis C virus; peptides; antibodies; ELISA.

XX Synthetic.

PH Key Location/Qualifiers
 FT Peptide 1..20 /label= 1
 FT Peptide 7..26 /label= 2
 FT Peptide 13..32 /label= 3
 FT Peptide 37..56 /label= 4
 FT Peptide 49..68 /label= 5
 FT Peptide 61..80

FT Peptide /label= 6
 FT 73...92
 FT /label= 7
 FT Peptide 1688..1707
 FT /label= 8
 FT Peptide 1694..1713
 FT /label= 9
 FT Peptide 1706..1725
 FT /label= 10
 FT Peptide 1712..1731
 FT /label= 11
 FT Peptide 1718..1737
 FT /label= 12
 FT Peptide 1724..1743
 FT /label= 13
 FT Peptide 1730..1749
 FT /label= 14
 FT Peptide 2263..2282
 FT /label= 15
 FT Peptide 2275..2294
 FT /label= 16
 FT Peptide 2287..2306
 FT /label= 17
 FT Peptide 2299..2318
 FT /label= 18
 FT Peptide 2311..2330
 FT /label= 19

PN EP489968-A1.

XX 17-JUN-1992.

XX 14-DEC-1990; 90EP-0124241.

XX 14-DEC-1990; 90EP-0124241.

XX (INNO-) INNOGENETICS NV.

XX Deleys RJ, Maertens G, Pollet D, Van Heuverswyn H;

XX WPI; 1992-201383/25.

XX New synthetic peptides for detecting antibodies to hepatitis C
 FT virus - useful in e.g. ELISA assays, and for detection of HCV
 FT antigens or as immunogens

XX Disclosure; Fig 1; 32pp; English.

CC RNA viruses frequently exhibit a high rate of spontaneous mutation,
 CC thus a virus is considered to be the same of equiv. to HCV if it
 CC exhibits a global homology of more than 70 percent with the HCV HC-
 CC J1/CDC/CHI composite sequence. The peptide fragments of this DNA
 CC sequence indicated in the features table can immunologically mimic
 CC proteins encoded by HCV. Additional amino acids or chemical gps.
 CC may be added to either end of the peptides for the purpose of
 CC creating a linker arm for attachment to a carrier. The peptides can
 CC be used for the detection of antibodies specific for HCV. They may
 CC be used in the form of kits, opt. with reagents such as
 CC staphylococcal protein A, streptococcal protein G, avidin or
 CC streptavidin. The peptides may also be used as immunogens for
 CC raising antibodies.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 2894 AA;

Query Match 100.0%; Score 20; DB 13; Length 2894;

Best Local Similarity 100.0%; Pred. No. 1.5e-11;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKNIQLINT 20

Db 397 SGLVSLFTPGAKNIQLINT 416

RESULT 14

AAR70230

XX AAR70230 standard; protein; 2894 AA.

XX AC AAR70230;

XX 25-MAR-2003 (updated)

XX 07-NOV-1995 (first entry)

XX Composite hepatitis C virus (HC-J1/CDC/CHI).

XX Composite hepatitis C virus; HC-J1/CDC/CHI; HCV; non-A non-B;
 XX synthetic antigens; blood screening.

XX Hepatitis C virus.

XX EP644202-A1.

XX 22-MAR-1995.

XX 14-DEC-1990; 94EP-0108611.

XX 14-DEC-1990; 90EP-0124241.

XX 14-DEC-1990; 90EP-0108611.

XX (INNO-) INNOGENETICS NV.

XX Deleys RJ, Maertens G, Pollet D, Van Heuverswyn H;

XX WPI; 1995-116946/16.

XX Synthetic antigens for the detection of hepatitis C virus
 FT antibodies - comprise portions of the HCV peptide sequence, for
 FT use in screening blood and blood products

XX Disclosure; Fig 1; 51pp; English.

CC AAR70230 is the composite hepatitis C virus (HC-J1/CDC/CHI) protein
 CC from which the synthetic HCV antigens described in AAR70210-R70229
 CC were derived. These synthetic antigens can be used to screen blood,
 CC or blood products for the presence of HCV, they can also be used in
 CC various specific assays for the detection of HCV antibodies, and
 CC antigens, or as immunogens.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC (Updated on 25-MAR-2003 to correct PF field.)

XX Sequence 2894 AA;

Query Match

Best Local Similarity 100.0%; Score 20; DB 16; Length 2894;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKNIQLINT 20

Db 397 SGLVSLFTPGAKNIQLINT 416

RESULT 15

AAR66995

XX AAR66995 standard; Protein; 3011 AA.

XX AC AAR66995;

XX 01-AUG-1995 (first entry)

XX Hepatitis C virus gene HC-J1/protein.

XX Hepatitis C virus; HCV gene HC-J1/protein; specific antibodies.

XX Hepatitis c virus.

XX JP06284887-A.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.55 Seconds
(without alignments)
117.536 Million cell updates/sec

Title: US-09-973-025-72

Perfect score: 20

Sequence: 1 SGLVSLFTPGAKQNIQLINT 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents AA New.*

- 1: /cgm2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
- 2: /cgm2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
- 3: /cgm2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 4: /cgm2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 5: /cgm2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 6: /cgm2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
- 7: /cgm2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	6	US-10-651-165-46
2	20	100.0	450	6	US-10-651-165-181
3	18	90.0	31	4	US-08-823-980F-69
4	13	65.0	30	6	US-10-296-734-460
5	13	65.0	30	6	US-10-296-734-462
6	13	65.0	450	6	US-10-651-165-179
7	13	65.0	1997	6	US-10-296-734-816
8	13	65.0	2010	6	US-10-296-734-814
9	13	65.0	3011	6	US-10-296-734-406
10	13	65.0	5985	6	US-10-296-734-810
11	11	55.0	31	4	US-08-823-980F-54
12	11	55.0	31	4	US-08-823-980F-59
13	9	45.0	31	4	US-08-823-980F-14
14	9	45.0	31	4	US-08-823-980F-51
15	9	45.0	31	4	US-08-823-980F-53
16	9	45.0	31	4	US-08-823-980F-57
17	9	45.0	31	4	US-08-823-980F-61
18	9	45.0	31	4	US-08-823-980F-74
19	9	45.0	31	4	US-08-823-980F-81
20	9	45.0	31	4	US-08-823-980F-82
21	8	40.0	20	6	US-10-651-165-47
22	7	35.0	31	4	US-08-823-980F-28
23	7	35.0	31	4	US-08-823-980F-32
24	7	35.0	31	4	US-08-823-980F-35
25	7	35.0	31	4	US-08-823-980F-38
26	7	35.0	31	4	US-08-823-980F-41

27	7	35.0	31	4	US-08-823-980F-52	Sequence 52, Appl
28	7	35.0	31	4	US-08-823-980F-56	Sequence 56, Appl
29	7	35.0	31	4	US-08-823-980F-62	Sequence 62, Appl
30	7	35.0	31	4	US-08-823-980F-97	Sequence 97, Appl
31	7	35.0	31	4	US-08-823-980F-100	Sequence 100, Appl
32	7	35.0	450	6	US-10-651-165-190	Sequence 190, Appl
33	6	30.0	11	6	US-10-440-390-7	Sequence 7, Appl
34	6	30.0	21	1	PCT-US03-28071-7	Sequence 7, Appl
35	6	30.0	21	1	PCT-US03-33610-5	Sequence 5, Appl
36	6	30.0	21	6	US-10-650-369-6	Sequence 6, Appl
37	6	30.0	21	6	US-10-658-782-7	Sequence 7, Appl
38	6	30.0	27	6	US-10-415-899-86	Sequence 86, Appl
39	6	30.0	30	6	US-10-296-734-458	Sequence 458, Appl
40	6	30.0	31	4	US-08-823-980F-16	Sequence 16, Appl
41	6	30.0	31	4	US-08-823-980F-19	Sequence 19, Appl
42	6	30.0	31	4	US-08-823-980F-21	Sequence 21, Appl
43	6	30.0	31	4	US-08-823-980F-29	Sequence 29, Appl
44	6	30.0	31	4	US-08-823-980F-30	Sequence 30, Appl
45	6	30.0	31	4	US-08-823-980F-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1
US-10-651-165-46
; Sequence 46, Application US/10651165
; GENERAL INFORMATION:

; APPLICANT: LEROUX-ROELS, Geert

; APPLICANT: DELEYS, Robert

; APPLICANT: MAERTENS, Geert

; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C

; TITLE OF INVENTION: VIRUS

; FILE REFERENCE: 2551-94

; CURRENT APPLICATION NUMBER: US/10/651,165

; CURRENT FILING DATE: 2003-09-02

; PRIOR APPLICATION NUMBER: US/08/974,690C

; PRIOR FILING DATE: 1997-11-19

; PRIOR APPLICATION NUMBER: PCT/EP94/03555

; PRIOR FILING DATE: 1994-10-28

; PRIOR APPLICATION NUMBER: EP 93402718.6

; PRIOR FILING DATE: 1993-11-04

; NUMBER OF SEQ ID NOS: 286

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 46

; LENGTH: 20

; TYPE: PRT

; ORGANISM: hepatitis C virus

US-10-651-165-46

Query Match 100.0%; Score 20; DB 6; Length 20;

Best Local Similarity 100.0%; Pred.No. 2.2e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGLVSLFTPGAKQNIQLINT 20

Db 1 SGLVSLFTPGAKQNIQLINT 20

RESULT 2

US-10-651-165-181

; Sequence 181, Application US/10651165

; GENERAL INFORMATION:

; APPLICANT: LEROUX-ROELS, Geert

; APPLICANT: DELEYS, Robert

; APPLICANT: MAERTENS, Geert

; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C

; TITLE OF INVENTION: VIRUS

; FILE REFERENCE: 2551-94

; CURRENT APPLICATION NUMBER: US/10/651,165

; CURRENT FILING DATE: 2003-09-02

; PRIOR APPLICATION NUMBER: US/08/974,690C

; PRIOR FILING DATE: 1997-11-19

; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 181
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-181

Query Match 100.0%; Score 20; DB 6; Length 450;
Best Local Similarity 100.0%; Pred. No. 1.9e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLINT 20
|||
Db 397 SGLVSLFTPGAKQNIQLINT 416
|||

RESULT 3

US-08-823-980F-69
; Sequence 69, Application US/08823980F
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; Houghton, Michael
; TITLE OF INVENTION: Conserved Motif of Hepatitis C Virus
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/823,980F
; FILING DATE: 25-Mar-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: CHIR-0108
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-325-7812
; TELEFAX: 650-325-7823

; INFORMATION FOR SEQ ID NO: 69:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 69:
US-08-823-980F-69

Query Match 90.0%; Score 18; DB 4; Length 31;
Best Local Similarity 100.0%; Pred. No. 2.6e-11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLI 18
|||
Db 14 SGLVSLFTPGAKQNIQLI 31
|||

RESULT 4

US-10-296-734-460
; Sequence 460, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 460
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 27
US-10-296-734-460

Query Match 65.0%; Score 13; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.8e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 TPGAKQNIQLINT 20
|||
Db 16 TPGAKQNIQLINT 28
|||

RESULT 5

US-10-296-734-462
; Sequence 462, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 462
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 28
US-10-296-734-462

Query Match 65.0%; Score 13; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.8e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 TPGAKQNIQLINT 20
|||
Db 1 TPGAKQNIQLINT 13
|||

RESULT 6

US-10-651-165-179
; Sequence 179, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02

```
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 179
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-179

Query Match      65.0%; Score 13; DB 6; Length 450;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      8 TPGAKQNIQLINT 20
Db      404 TPGAKQNIQLINT 416

RESULT 7
US-10-296-734-816
; Sequence 816, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 816
; LENGTH: 1997
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette C
US-10-296-734-816

Query Match      65.0%; Score 13; DB 6; Length 1997;
Best Local Similarity 100.0%; Pred. No. 3.4e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      8 TPGAKQNIQLINT 20
Db      571 TPGAKQNIQLINT 583

RESULT 8
US-10-296-734-814
; Sequence 814, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 814
; LENGTH: 2010
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette B
US-10-296-734-814

Query Match      65.0%; Score 13; DB 6; Length 2010;
Best Local Similarity 100.0%; Pred. No. 3.5e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      8 TPGAKQNIQLINT 20
Db      166 TPGAKQNIQLINT 178

RESULT 9
US-10-296-734-406
; Sequence 406, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a consensus polyprotein
US-10-296-734-406

Query Match      65.0%; Score 13; DB 6; Length 3011;
Best Local Similarity 100.0%; Pred. No. 4.6e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      8 TPGAKQNIQLINT 20
Db      404 TPGAKQNIQLINT 416

RESULT 10
US-10-296-734-810
; Sequence 810, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 810
; LENGTH: 5985
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a savine
US-10-296-734-810

Query Match      65.0%; Score 13; DB 6; Length 5985;
Best Local Similarity 100.0%; Pred. No. 7.4e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      8 TPGAKQNIQLINT 20
Db      2176 TPGAKQNIQLINT 2188
```

```
RESULT 11
US-08-823-980F-54
; Sequence 54, Application US/08823980F
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; Houghton, Michael
; TITLE OF INVENTION: Conserved Motif of Hepatitis C Virus
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/823,980F
; FILING DATE: 25-Mar-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: CHIR-0108
; TELEPHONE: 650-325-7812
; TELEFAX: 650-325-7823
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 54:
US-08-823-980F-54

Query Match 55.0%; Score 11; DB 4; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 TPGAKQNIQLI 18
Db 21 TPGAKQNIQLI 31

RESULT 12
US-08-823-980F-99
; Sequence 99, Application US/08823980F
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; Houghton, Michael
; TITLE OF INVENTION: Conserved Motif of Hepatitis C Virus
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/823,980F
; FILING DATE: 25-Mar-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: CHIR-0108
; TELEPHONE: 650-325-7812
; TELEFAX: 650-325-7823
; INFORMATION FOR SEQ ID NO: 99:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 99:
US-08-823-980F-99

Query Match 45.0%; Score 9; DB 4; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 TPGAKQNIQLI 18
Db 21 TPGAKQNIQLI 31
```

```
US-08-823-980F-14
; Sequence 14, Application US/08823980F
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; Houghton, Michael
; TITLE OF INVENTION: Conserved Motif of Hepatitis C Virus
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/823,980F
; FILING DATE: 25-Mar-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: CHIR-0108
; TELEPHONE: 650-325-7812
; TELEFAX: 650-325-7823
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-08-823-980F-14

Query Match 55.0%; Score 11; DB 4; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 TPGAKQNIQLI 18
Db 21 TPGAKQNIQLI 31

RESULT 13
US-08-823-980F-14
; Sequence 14, Application US/08823980F
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; Houghton, Michael
; TITLE OF INVENTION: Conserved Motif of Hepatitis C Virus
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/823,980F
; FILING DATE: 25-Mar-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: CHIR-0108
; TELEPHONE: 650-325-7812
; TELEFAX: 650-325-7823
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-08-823-980F-14

Query Match 45.0%; Score 9; DB 4; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 TPGAKQNIQLI 18
Db 21 TPGAKQNIQLI 31
```


Best Local Similarity 100.0%; Pred. No. 0.015; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

Oy 10 GAKONIQLI 18
| | | | | | | |
Db 23 GAKONIQLI 31

RESULT 14

US-08-823-980F-51

; Sequence 51, Application US/08823980F

; GENERAL INFORMATION:

; APPLICANT: Weiner, Amy J.

; Houghton, Michael

; TITLE OF INVENTION: Conserved Motif of Hepatitis C Virus

; E2/NS1 Region

; NUMBER OF SEQUENCES: 134

; CORRESPONDENCE ADDRESS: 134

; ADDRESSEE: Chiron Corporation

; STREET: 4560 Horton Street

; CITY: Emeryville

; STATE: CA

; COUNTRY: USA

; ZIP: 94608-2916

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/823,980F

; FILING DATE: 25-Mar-1997

; CLASSIFICATION: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Robins, Roberta L.

; REGISTRATION NUMBER: 33,208

; REFERENCE/DOCKET NUMBER: CHIR-0108

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-325-7812

; TELEFAX: 650-325-7823

; INFORMATION FOR SEQ ID NO: 51:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 31 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; SEQUENCE DESCRIPTION: SEQ ID NO: 51:

US-08-823-980F-51

Query Match 45.0%; Score 9; DB 4; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.015; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

Oy 10 GAKONIQLI 18
| | | | | | | |
Db 23 GAKONIQLI 31

RESULT 15

US-08-823-980F-53

; Sequence 53, Application US/08823980F

; GENERAL INFORMATION:

; APPLICANT: Weiner, Amy J.

; Houghton, Michael

; TITLE OF INVENTION: Conserved Motif of Hepatitis C Virus

; E2/NS1 Region

; NUMBER OF SEQUENCES: 134

; CORRESPONDENCE ADDRESS: 134

; ADDRESSEE: Chiron Corporation

; STREET: 4560 Horton Street

; CITY: Emeryville

; STATE: CA

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time 19.5 Seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-72

Perfect score: 20

Sequence: 1 SGLVSLTPGAKQNIQLINT 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	10	US-09-973-025-72
2	20	100.0	20	11	US-09-899-303-72
3	20	100.0	20	11	US-09-995-808-72
4	20	100.0	20	11	US-09-995-860-72
5	20	100.0	20	12	US-09-995-791-72
6	20	100.0	23	16	US-10-318-200-26
7	20	100.0	2894	10	US-09-941-611-23
8	20	100.0	2894	15	US-10-044-995-23
9	13	65.0	91	10	US-09-921-397-80
10	13	65.0	160	12	US-10-292-129-15
11	13	65.0	350	10	US-09-929-955-4
12	13	65.0	350	14	US-10-104-966-4
13	13	65.0	363	12	US-10-128-587A-97
14	13	65.0	363	15	US-10-128-590-97
15	13	65.0	3011	9	US-09-742-659-4

16	13	65.0	3011	10	US-09-238-076-20	Sequence 20, Appl
17	13	65.0	3011	10	US-09-952-572-9	Sequence 9, Appl
18	13	65.0	3011	10	US-09-929-955-1	Sequence 1, Appl
19	13	65.0	3011	10	US-09-747-419-20	Sequence 20, Appl
20	13	65.0	3011	11	US-09-891-894-3	Sequence 3, Appl
21	13	65.0	3011	11	US-09-995-937-20	Sequence 20, Appl
22	13	65.0	3011	11	US-09-917-563-20	Sequence 20, Appl
23	13	65.0	3011	12	US-10-184-150-3	Sequence 3, Appl
24	13	65.0	3011	14	US-10-104-966-1	Sequence 1, Appl
25	13	65.0	3011	15	US-10-259-275-20	Sequence 20, Appl
26	13	65.0	3012	10	US-09-238-076-2	Sequence 2, Appl
27	13	65.0	3012	11	US-09-995-937-2	Sequence 2, Appl
28	13	65.0	3012	11	US-09-917-563-2	Sequence 2, Appl
29	11	55.0	31	15	US-10-196-174-2	Sequence 2, Appl
30	8	40.0	20	10	US-09-973-025-73	Sequence 73, Appl
31	8	40.0	20	11	US-09-899-303-73	Sequence 73, Appl
32	8	40.0	20	11	US-09-995-808-73	Sequence 73, Appl
33	8	40.0	20	11	US-09-995-860-73	Sequence 73, Appl
34	8	40.0	20	12	US-09-995-791-73	Sequence 73, Appl
35	7	35.0	62	10	US-09-764-868-1132	Sequence 1132, Ap
36	6	30.0	10	8	US-08-854-825-5	Sequence 5, Appl
37	6	30.0	21	10	US-09-881-654-5	Sequence 5, Appl
38	6	30.0	21	10	US-09-881-239-6	Sequence 6, Appl
39	6	30.0	21	12	US-10-265-083-4	Sequence 4, Appl
40	6	30.0	21	16	US-10-318-200-31	Sequence 31, Appl
41	6	30.0	23	16	US-10-318-200-23	Sequence 23, Appl
42	6	30.0	23	16	US-10-318-200-28	Sequence 28, Appl
43	6	30.0	24	16	US-10-318-200-22	Sequence 22, Appl
44	6	30.0	31	15	US-10-196-174-1	Sequence 1, Appl
45	6	30.0	31	15	US-10-196-174-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1

US-09-973-025-72
; Sequence 72, Application US/09973025
; Publication No. US20020182706A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 72:
US-09-973-025-72

Query Match 100.0%; Score 20; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLINT 20
Db 1 SGLVSLFTPGAKQNIQLINT 20
|||||

RESULT 2

US-09-899-303-72
; Sequence 72, Application US/09899303
; Publication No. US20030036110A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/899,303

FILING DATE: 06-Jul-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 72:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 72:

US-09-899-303-72

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLINT 20
|||||

Db 1 SGLVSLFTPGAKQNIQLINT 20

RESULT 3

US-09-995-808-72
; Sequence 72, Application US/09995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:

; APPLICANT: Innogenetics N.V.

; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and

; TITLE OF INVENTION: therapeutic use.

; FILE REFERENCE: 2551-70

; CURRENT APPLICATION NUMBER: US/09/995,808

; CURRENT FILING DATE: 2001-11-29

; NUMBER OF SEQ ID NOS: 122

; SOFTWARE: PatentIn 3.1

; SEQ ID NO 72

; LENGTH: 20

; TYPE: PRT

; ORGANISM: Hepatitis C virus

US-09-995-808-72

Query Match 100.0%; Score 20; DB 11; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.2e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLINT 20
|||||

Db 1 SGLVSLFTPGAKQNIQLINT 20
|||||

RESULT 4

US-09-995-860-72

; Sequence 72, Application US/09995860

; Publication No. US20030118603A1

; GENERAL INFORMATION:

; APPLICANT: Innogenetics N.V.

; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and

; TITLE OF INVENTION: therapeutic use.

; FILE REFERENCE: 2551-69

; CURRENT APPLICATION NUMBER: US/09/995,860

; CURRENT FILING DATE: 2001-11-29

; NUMBER OF SEQ ID NOS: 122

; SOFTWARE: PatentIn 3.1

; SEQ ID NO 72

; LENGTH: 20

; TYPE: PRT

; ORGANISM: Hepatitis C virus

US-09-995-860-72

Query Match 100.0%; Score 20; DB 11; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.2e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLINT 20
|||||

Db 1 SGLVSLFTPGAKQNIQLINT 20
|||||

RESULT 5

US-09-995-791-72

; Sequence 72, Application US/09995791

; Publication No. US20030147918A1

; GENERAL INFORMATION:

; APPLICANT: Innogenetics N.V.

; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and

; TITLE OF INVENTION: therapeutic use.

; FILE REFERENCE: 2551-68

; CURRENT APPLICATION NUMBER: US/09/995,791

; CURRENT FILING DATE: 2001-11-29

; NUMBER OF SEQ ID NOS: 122

; SOFTWARE: PatentIn 3.1

; SEQ ID NO 72

```
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-72

Query Match      100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 SGLVSLFTPGAKONIQLINT 20
      |||||
Db      1 SGLVSLFTPGAKONIQLINT 20

RESULT 6
US-10-318-200-26
; Sequence 26, Application US/10318200
; Publication No. US20030129746A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BUYSE, ERIC
; TITLE OF INVENTION: EPITOPES IN VIRAL ENVELOPE PROTEINS AND SPECIFIC
; TITLE OF INVENTION: ANTIBODIES DIRECTED AGAINST THESE EPITOPES: USE FOR
; TITLE OF INVENTION: DETECTION OF HCV VIRAL ANTIGEN IN HOST TISSUE
; FILE REFERENCE: 2551-47
; CURRENT APPLICATION NUMBER: US/10/318,200
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: US/09/645,470
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: PCT/EP99/02154
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: EP 98870060.5
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-318-200-26

Query Match      100.0%; Score 20; DB 16; Length 23;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 SGLVSLFTPGAKONIQLINT 20
      |||||
Db      4 SGLVSLFTPGAKONIQLINT 23

RESULT 7
US-09-941-611-23
; Sequence 23, Application US/09941611
; Patent No. US20020106640A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; TITLE OF INVENTION: ANTIBODIES TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

US-09-941-611-23
; Sequence 23, Application US/10044995
; Publication No. US20030049685A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; TITLE OF INVENTION: ANTIBODIES TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/941,611
; FILING DATE: 30-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/391,671
; FILING DATE: 1995-02-21
; APPLICATION NUMBER: WO PCT/EP91/02409
; FILING DATE: 13-DEC-1991
; APPLICATION NUMBER: EP 90124241.2
; FILING DATE: 14-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 1487-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 7038164000
; TELEFAX: 7038164100
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-941-611-23

Query Match      100.0%; Score 20; DB 10; Length 2894;
Best Local Similarity 100.0%; Pred. No. 8.1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 SGLVSLFTPGAKONIQLINT 20
      |||||
Db      397 SGLVSLFTPGAKONIQLINT 416

RESULT 8
US-10-044-995-23
; Sequence 23, Application US/10044995
; Publication No. US20030049685A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; TITLE OF INVENTION: ANTIBODIES TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/044,995
; FILING DATE: 15-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/391,671
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 07/920,286
```

```
; FILING DATE: 14-OCT-1992
; APPLICATION NUMBER: WO PCT/EP91/02409
; FILING DATE: 13-DEC-1991
; APPLICATION NUMBER: EP 90124241.2
; FILING DATE: 14-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36.663
; REFERENCE/DOCKET NUMBER: 1487-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 7038164000
; TELEFAX: 7038164100
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-044-995-23

Query Match          100.0%; Score 20; DB 15; Length 2894;
Best Local Similarity 100.0%; Pred. No. 8.1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGLVSLFTPGAKQNIQLINT 20
DB 397 SGLVSLFTPGAKQNIQLINT 416

RESULT 9
US-09-921-397-80
; Sequence 80, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 80
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-80

Query Match          65.0%; Score 13; DB 10; Length 91;
Best Local Similarity 100.0%; Pred. No. 5.2e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TPGAKQNIQLINT 20
DB 25 TPGAKQNIQLINT 37

RESULT 10
US-10-129-129-15
; Sequence 15, Application US/10292129
; Publication No. US20030148267A1
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Emmett Vance
; APPLICANT: Chung, Raymond Taeyong
; TITLE OF INVENTION: SCREENING ASSAY FOR HEPATITIS C VIRUS
; TITLE OF INVENTION: ANTIVIRAL AGENTS
```

```
; FILE REFERENCE: 00786-539001
; CURRENT APPLICATION NUMBER: US/10/292,129
; CURRENT FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: US 60/345,405
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-292-129-15

Query Match          65.0%; Score 13; DB 12; Length 160;
Best Local Similarity 100.0%; Pred. No. 8.4e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TPGAKQNIQLINT 20
DB 136 TPGAKQNIQLINT 148

RESULT 11
US-09-929-955-4
; Sequence 4, Application US/09929955
; Patent No. US20020136740A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.23AUS2
; CURRENT APPLICATION NUMBER: US/09/929,955
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/225,767
; PRIOR FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE: c
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-09-929-955-4

Query Match          65.0%; Score 13; DB 10; Length 350;
Best Local Similarity 100.0%; Pred. No. 1.6e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TPGAKQNIQLINT 20
DB 25 TPGAKQNIQLINT 37

RESULT 12
US-10-104-966-4
; Sequence 4, Application US/10104966
; Publication No. US20020155124A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.23AUSC1
; CURRENT APPLICATION NUMBER: US/10/104,966
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
```

; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-10-104-966-4

Query Match 65.0%; Score 13; DB 14; Length 350;
Best Local Similarity 100.0%; Pred.No. 1.6e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 TPGAKQNIQLINT 20
Db 25 TPGAKQNIQLINT 37
|||||

RESULT 13

US-10-128-587A-97
; Sequence 97, Application US/10128587A
; Publication No. US20030152940A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Constructs and methods for expression of recombinant HCV envelope

; FILE REFERENCE: 134 PCT
; CURRENT APPLICATION NUMBER: US/10/128,587A
; CURRENT FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: hepatitis C virus
US-10-128-587A-97

Query Match 65.0%; Score 13; DB 12; Length 363;
Best Local Similarity 100.0%; Pred.No. 1.7e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 TPGAKQNIQLINT 20
Db 21 TPGAKQNIQLINT 33
|||||

RESULT 14

US-10-128-590-97
; Sequence 97, Application US/10128590
; Publication No. US20030108561A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Core-glycosylated HCV envelope proteins

; FILE REFERENCE: 135 PCT
; CURRENT APPLICATION NUMBER: US/10/128,590
; CURRENT FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-128-590-97

Query Match 65.0%; Score 13; DB 15; Length 363;
Best Local Similarity 100.0%; Pred.No. 1.7e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 TPGAKQNIQLINT 20
Db 21 TPGAKQNIQLINT 33
|||||

RESULT 15

US-09-742-659-4
; Sequence 4, Application US/09742659
; Patent No. US20010034019A1
; GENERAL INFORMATION:
; APPLICANT: Hong, Zhi

; APPLICANT: Buckiewicz, Nancy J.
; APPLICANT: Zhong, Weidong
; APPLICANT: Ingravallo, Paul
; APPLICANT: Wright-Minogue, Jacquelyn
; APPLICANT: Lau, Johnson Y.
; APPLICANT: Lemon, Stanley M.
; TITLE OF INVENTION: Chimeric HCV/GBV-B viruses
; FILE REFERENCE: ID01116
; CURRENT APPLICATION NUMBER: US/09/742,659
; CURRENT FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: US 60/171,469
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-742-659-4

Query Match 65.0%; Score 13; DB 9; Length 3011;
Best Local Similarity 100.0%; Pred.No. 0.0001;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 TPGAKQNIQLINT 20
Db 404 TPGAKQNIQLINT 416
|||||

Search completed: November 21, 2003, 22:19:36
Job time : 20.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 163.1 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-72

Perfect score: 20

Sequence: 1 SGLVSLTPGAKQNIQLINT 20

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA_Main.*

- 1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/1/paa/US097A_COMB.pep.*
- 22: /cgn2_6/ptodata/1/paa/US097B_COMB.pep.*
- 23: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
- 24: /cgn2_6/ptodata/1/paa/US099A_COMB.pep.*
- 25: /cgn2_6/ptodata/1/paa/US099B_COMB.pep.*
- 26: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
- 27: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
- 28: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
- 29: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
- 30: /cgn2_6/ptodata/1/paa/US104_COMB.pep.*
- 31: /cgn2_6/ptodata/1/paa/US106_COMB.pep.*
- 32: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	20	100.0	20	13	US-08-974-685-46
					Sequence 46, Appl

2	20	100.0	20	13	US-08-974-690-46	Sequence 46, Appl
3	20	100.0	20	13	US-08-974-690A-46	Sequence 46, Appl
4	20	100.0	20	13	US-08-974-690B-46	Sequence 46, Appl
5	20	100.0	20	13	US-08-974-690C-46	Sequence 46, Appl
6	20	100.0	20	23	US-09-899-303-72	Sequence 72, Appl
7	20	100.0	20	23	US-09-899-303A-72	Sequence 72, Appl
8	20	100.0	20	25	US-09-973-025-72	Sequence 72, Appl
9	20	100.0	20	25	US-09-995-791-72	Sequence 72, Appl
10	20	100.0	20	25	US-09-995-808-72	Sequence 72, Appl
11	20	100.0	20	25	US-09-995-860-72	Sequence 72, Appl
12	20	100.0	20	26	US-10-020-510-72	Sequence 72, Appl
13	20	100.0	20	29	US-10-321-798-72	Sequence 72, Appl
14	20	100.0	23	29	US-10-318-200-26	Sequence 26, Appl
15	20	100.0	24	19	US-09-576-824A-419	Sequence 419, App
16	20	100.0	24	22	US-09-790-497A-85	Sequence 85, Appl
17	20	100.0	24	22	US-09-790-497A-85	Sequence 85, Appl
18	20	100.0	24	22	US-09-790-497A-419	Sequence 419, App
19	20	100.0	24	22	US-09-790-497A-474	Sequence 474, App
20	20	100.0	24	22	US-09-790-812A-85	Sequence 85, Appl
21	20	100.0	24	22	US-09-790-812A-419	Sequence 419, App
22	20	100.0	24	22	US-09-790-812A-474	Sequence 474, App
23	20	100.0	24	31	US-10-621-675-419	Sequence 419, App
24	20	100.0	26	11	US-08-723-425-85	Sequence 85, Appl
25	20	100.0	26	19	US-09-576-824-85	Sequence 85, Appl
26	20	100.0	26	19	US-09-576-824A-474	Sequence 474, App
27	20	100.0	26	19	US-09-576-824A-85	Sequence 85, Appl
28	20	100.0	26	19	US-09-576-824A-474	Sequence 474, App
29	20	100.0	26	20	US-09-680-497-85	Sequence 85, Appl
30	20	100.0	26	31	US-10-621-675-85	Sequence 85, Appl
31	20	100.0	26	31	US-10-621-675-474	Sequence 474, App
32	20	100.0	34	19	US-09-576-824-131	Sequence 131, App
33	20	100.0	34	19	US-09-576-824A-131	Sequence 131, App
34	20	100.0	34	22	US-09-790-497A-83	Sequence 83, Appl
35	20	100.0	34	22	US-09-790-497A-131	Sequence 131, App
36	20	100.0	34	22	US-09-790-497A-459	Sequence 459, App
37	20	100.0	34	22	US-09-790-812A-83	Sequence 83, Appl
38	20	100.0	34	22	US-09-790-812A-131	Sequence 131, App
39	20	100.0	34	22	US-09-790-812A-459	Sequence 459, App
40	20	100.0	34	31	US-10-621-675-131	Sequence 131, App
41	20	100.0	36	11	US-08-723-425-83	Sequence 83, Appl
42	20	100.0	36	11	US-08-723-425-131	Sequence 131, App
43	20	100.0	36	19	US-09-576-824-83	Sequence 83, Appl
44	20	100.0	36	19	US-09-576-824A-459	Sequence 459, App
45	20	100.0	36	19	US-09-576-824A-83	Sequence 83, Appl

ALIGNMENTS

RESULT 1

US-08-974-685-46
; Sequence 46, Application US/08974685
; GENERAL INFORMATION:
; APPLICANT: LEROUIX-ROELS, GEERT
; DELEYS, ROBERT
; MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; HEPATITIS C VIRUS STRUCTURAL PROTEINS. PROPHYLACTIC AND
; THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 181
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,685
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-08-974-685-46

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLINT 20
Db 1 SGLVSLFTPGAKQNIQLINT 20

RESULT 2

US-08-974-690-46
; Sequence 46, Application US/08974690
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; APPLICANT: DELEYS, ROBERT
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; TITLE OF INVENTION: HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
; TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 166
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION NUMBER: US/08/974,690
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/635,886
; FILING DATE: 25-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

US-08-974-690-46

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLINT 20
Db 1 SGLVSLFTPGAKQNIQLINT 20

RESULT 3

US-08-974-690A-46
; Sequence 46, Application US/08974690A
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; APPLICANT: DELEYS, ROBERT
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; TITLE OF INVENTION: HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
; TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION NUMBER: US/08/974,690A
; FILING DATE: 19-Nov-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B. J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 2752-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-08-974-690A-46

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLINT 20
Db 1 SGLVSLFTPGAKQNIQLINT 20

RESULT 4

US-08-974-690B-46
; Sequence 46, Application US/08974690B
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; APPLICANT: DELEYS, ROBERT
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; TITLE OF INVENTION: HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690B
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-08-974-690B-46

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLINT 20
Db 1 SGLVSLFTPGAKQNIQLINT 20

RESULT 5
US-08-974-690C-46
SEQUENCE 46, Application US/08974690C
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELEYS, Robert
APPLICANT: MAERTENS, Geert
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
FILE REFERENCE: 2551-94
CURRENT APPLICATION NUMBER: US/08/974,690C
CURRENT FILING DATE: 1997-11-19
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28
PRIOR APPLICATION NUMBER: EP 93402718.6
PRIOR FILING DATE: 1993-11-04
NUMBER OF SEQ ID NOS: 286
SOFTWARE: PatentIn version 3.1
SEQ ID NO 46
LENGTH: 20
TYPE: PRT
ORGANISM: hepatitis C virus
US-08-974-690C-46

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLINT 20

Db 1 SGLVSLFTPGAKQNIQLINT 20

RESULT 6
US-09-899-303-72
SEQUENCE 72, Application US/09899303
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 72:
US-09-899-303-72

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLINT 20
Db 1 SGLVSLFTPGAKQNIQLINT 20

RESULT 7
US-09-899-303A-72
SEQUENCE 72, Application US/09899303A
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.

STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303A
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 72:
US-09-899-303A-72
Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SGLVSLFTPGAKQNIQLINT 20
Db 1 SGLVSLFTPGAKQNIQLINT 20
RESULT 8
US-09-973-025-72
; Sequence 72, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; BUYSE, MARIE-ANGE
; DE MARTYNOFF, GUY
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 72:
US-09-973-025-72
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SGLVSLFTPGAKQNIQLINT 20
Db 1 SGLVSLFTPGAKQNIQLINT 20
RESULT 9
US-09-995-791-72
; Sequence 72, Application US/09995791
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; therapeutic use.
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 72
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; US-09-995-791-72
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SGLVSLFTPGAKQNIQLINT 20
Db 1 SGLVSLFTPGAKQNIQLINT 20
RESULT 10
US-09-995-808-72
; Sequence 72, Application US/09995808
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 72
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; US-09-995-808-72
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SGLVSLFTPGAKQNIQLINT 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 SGLVSLFTPGAKQNIQLINT 20
| | | | | | | | | | | | | | | | | | | | | |

RESULT 11
US-09-995-860-72
; Sequence 72, Application US/09995860
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 72
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-72

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SGLVSLFTPGAKQNIQLINT 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 SGLVSLFTPGAKQNIQLINT 20
| | | | | | | | | | | | | | | | | | | | | |

RESULT 12
US-10-020-510-72
; Sequence 72, Application US/10020510
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-72
; CURRENT APPLICATION NUMBER: US/10/020,510
; CURRENT FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 72
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-020-510-72

Query Match 100.0%; Score 20; DB 26; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SGLVSLFTPGAKQNIQLINT 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 SGLVSLFTPGAKQNIQLINT 20
| | | | | | | | | | | | | | | | | | | | | |

RESULT 13
US-10-321-798-72
; Sequence 72, Application US/10321798
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-93
; CURRENT APPLICATION NUMBER: US/10/321,798
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: 60/418,358
; PRIOR FILING DATE: 2002-10-16
```

```
; PRIOR APPLICATION NUMBER: 10/020,510
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 72
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-321-798-72

Query Match 100.0%; Score 20; DB 29; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SGLVSLFTPGAKQNIQLINT 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 SGLVSLFTPGAKQNIQLINT 20
| | | | | | | | | | | | | | | | | | | | | |

RESULT 14
US-10-318-200-26
; Sequence 26, Application US/10318200
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: DELPA, ERIK
; TITLE OF INVENTION: EPITOPES IN VIRAL ENVELOPE PROTEINS AND SPECIFIC
; TITLE OF INVENTION: ANTIBODIES DIRECTED AGAINST THESE EPITOPES: USE FOR
; TITLE OF INVENTION: DETECTION OF HCV VIRAL ANTIGEN IN HOST TISSUE
; FILE REFERENCE: 2551-47
; CURRENT APPLICATION NUMBER: US/10/318,200
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: US/09/645,470
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: PCT/EP99/02154
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: EP 98870060.5
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-318-200-26

Query Match 100.0%; Score 20; DB 29; Length 23;
Best Local Similarity 100.0%; Pred. No. 4.9e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SGLVSLFTPGAKQNIQLINT 20
| | | | | | | | | | | | | | | | | | | | | |
Db 4 SGLVSLFTPGAKQNIQLINT 23
| | | | | | | | | | | | | | | | | | | | | |

RESULT 15
US-09-576-824-419
; Sequence 419, Application US/09576824
; GENERAL INFORMATION:
; APPLICANT: De Leys, Robert
; TITLE OF INVENTION: PROCESS FOR THE DETERMINATION OF PEPTIDES CORRESPONDING
; TITLE OF INVENTION: TO IMMUNOLOGICALLY IMPORTANT EPITOPES AND THEIR USE IN
; TITLE OF INVENTION: A PROCESS FOR DETERMINATION OF ANTIBODIES OF
; TITLE OF INVENTION: BIOTINYLATED PEPTIDES CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT
; TITLE OF INVENTION: EPITOPES, A PROCESS FOR PREPARING THEM AND COMPOSITIONS
; TITLE OF INVENTION: CONTAINING THEM
; FILE REFERENCE: 2752-11
; CURRENT APPLICATION NUMBER: US/09/576,824
; CURRENT FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 08/723,425
; PRIOR FILING DATE: 1996-09-30
; PRIOR APPLICATION NUMBER: PCT/EP93/00517
; PRIOR FILING DATE: 1993-03-08
```

```

; PRIOR APPLICATION NUMBER: EP 92400598.8
; PRIOR FILING DATE: 1992-03-06
; NUMBER OF SEQ ID NOS: 600
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 419
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-576-824-419

```

```
Query Match      100.0%; Score 20; DB 19; Length 24;
Best Local Similarity 100.0%; Pred. No. 5.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	1	SGLVSLFTPGAKQNIQLINT	20
Db	5	SGLVSLFTPGAKQNIQLINT	24

Search completed: November 21, 2003, 22:09:49
Job time : 164.1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.6 Seconds
(without alignments)
79,832 Million cell updates/sec

Title: US-09-973-025-72
Perfect score: 20
Sequence: 1 SGLVSLFPGAKQNIQLINT 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-08-612-973-72
2	20	100.0	20	3	US-08-927-597-72
3	20	100.0	20	4	US-08-635-886C-46
4	20	100.0	23	4	US-09-645-470-26
5	20	100.0	26	2	US-08-146-028-85
6	20	100.0	26	3	US-08-723-425A-85
7	20	100.0	26	3	US-09-112-206-85
8	20	100.0	36	2	US-08-146-028-83
9	20	100.0	36	3	US-08-146-028-131
10	20	100.0	36	3	US-08-723-425A-83
11	20	100.0	36	3	US-08-723-425A-131
12	20	100.0	36	3	US-09-112-206-83
13	20	100.0	36	3	US-09-112-206-131
14	20	100.0	144	1	US-08-440-103-20
15	20	100.0	144	1	US-08-440-542-20
16	20	100.0	144	1	US-08-231-368-20
17	20	100.0	144	1	US-08-440-210-20
18	20	100.0	144	1	US-09-046-604-20
19	20	100.0	174	1	US-08-460-806-19
20	20	100.0	174	1	US-08-325-630-19
21	20	100.0	180	2	US-08-483-695-40
22	20	100.0	180	2	US-07-965-285-40
23	20	100.0	180	2	US-08-487-231-40
24	20	100.0	180	3	US-09-201-912-40
25	20	100.0	450	4	US-08-635-886C-181
26	20	100.0	2894	2	US-08-466-975A-23
27	20	100.0	2894	2	US-08-391-671A-23

28	20	100.0	2894	3	US-08-467-902A-23	Sequence 23, Appl
29	20	100.0	2894	3	US-09-275-265-23	Sequence 23, Appl
30	20	100.0	2894	4	US-09-941-611-23	Sequence 23, Appl
31	19	95.0	25	1	US-08-262-037-87	Sequence 87, Appl
32	19	95.0	35	1	US-08-262-037-9	Sequence 9, Appl
33	19	95.0	45	1	US-08-262-037-88	Sequence 88, Appl
34	19	95.0	55	1	US-08-262-037-89	Sequence 89, Appl
35	19	95.0	60	1	US-08-262-037-90	Sequence 90, Appl
36	13	65.0	402	1	US-08-460-806-15	Sequence 15, Appl
37	13	65.0	402	1	US-08-325-630-15	Sequence 15, Appl
38	13	65.0	409	1	US-08-440-103-21	Sequence 21, Appl
39	13	65.0	409	1	US-08-440-103-24	Sequence 21, Appl
40	13	65.0	409	1	US-08-440-542-21	Sequence 21, Appl
41	13	65.0	409	1	US-08-440-542-24	Sequence 21, Appl
42	13	65.0	409	1	US-08-231-368-21	Sequence 21, Appl
43	13	65.0	409	1	US-08-231-368-24	Sequence 21, Appl
44	13	65.0	409	1	US-08-440-210-21	Sequence 21, Appl
45	13	65.0	409	1	US-08-440-210-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-08-612-973-72
; Sequence 72, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-612-973-72

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.le-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SGLVSLFPGAKQNIQLINT 20
|||||

Db 1 SGLVSLFTPGAKONIQLINT 20

RESULT 2

US-08-927-597-72
; Sequence 72, Application US/08927597
; Patent No. 6245503

; GENERAL INFORMATION:

; APPLICANT: MAERTENS, GEERT

; APPLICANT: BOSMAN, FONS

; APPLICANT: DE MARTYNOFF, GUY

; APPLICANT: BUYSE, MARIE-ANGE

; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

; NUMBER OF SEQUENCES: 111

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: NIXON & VANDERHUYE P.C.

; STREET: 1100 NORTH GLEBE ROAD

; CITY: ARLINGTON

; STATE: VIRGINIA

; COUNTRY: U.S.A.

; ZIP: 22201-4714

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/927,597

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/612,973

; FILING DATE: 11-MAR-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: BYRNE, THOMAS E.

; REGISTRATION NUMBER: 32,205

; REFERENCE/DOCKET NUMBER: 1487-10

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 816-4000

; TELEFAX: (703) 816-4100

; INFORMATION FOR SEQ ID NO: 72:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 20 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-927-597-72

Query Match 100.0%; Score 20; DB 3; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.1e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKONIQLINT 20

Db 1 SGLVSLFTPGAKONIQLINT 20

RESULT 3

US-08-635-886C-46

; Sequence 46, Application US/08635886C

; Patent No. 6555114

; GENERAL INFORMATION:

; APPLICANT: LEROUX-ROELS, Geert

; APPLICANT: DELEYS, Robert

; APPLICANT: MAERTENS, Geert

; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C

; FILE REFERENCE: 2752-18

; CURRENT APPLICATION NUMBER: US/08/635,886C

; CURRENT FILING DATE: 1996-04-25

; PRIOR APPLICATION NUMBER: PCT/EP94/03555

; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6

; PRIOR FILING DATE: 1993-11-04

; NUMBER OF SEQ ID NOS: 286

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 46

; LENGTH: 20

; TYPE: PRT

; ORGANISM: hepatitis C virus

US-08-635-886C-46

Query Match 100.0%; Score 20; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.1e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKONIQLINT 20

Db 1 SGLVSLFTPGAKONIQLINT 20

RESULT 4

US-09-645-470-26

; Sequence 26, Application US/09645470

; Patent No. 6521403

; GENERAL INFORMATION:

; APPLICANT: MAERTENS, GEERT

; APPLICANT: DELPA, ERIK

; APPLICANT: BUYSE, MARIE-ANGE

; TITLE OF INVENTION: EPITOPES IN VIRAL ENVELOPE PROTEINS AND SPECIFIC

; TITLE OF INVENTION: ANTIBODIES DIRECTED AGAINST THESE EPITOPES: USE FOR

; TITLE OF INVENTION: DETECTION OF HCV VIRAL ANTIGEN IN HOST TISSUE

; FILE REFERENCE: 2551-47

; CURRENT APPLICATION NUMBER: US/09/645,470

; CURRENT FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: PCT/EP99/02154

; PRIOR FILING DATE: 1999-03-29

; PRIOR APPLICATION NUMBER: EP 98870060.5

; PRIOR FILING DATE: 1998-03-27

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn ver. 2.1

; SEQ ID NO 26

; LENGTH: 23

; TYPE: PRT

; ORGANISM: Hepatitis C virus

US-09-645-470-26

Query Match 100.0%; Score 20; DB 4; Length 23;

Best Local Similarity 100.0%; Pred. No. 1.2e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKONIQLINT 20

Db 4 SGLVSLFTPGAKONIQLINT 23

RESULT 5

US-08-146-028-85

; Sequence 85, Application US/08146028

; Patent No. 5891640

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: PROCESS FOR THE DETERMINATION OF PEPTIDES

; TITLE OF INVENTION: CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES AND THEIR

; TITLE OF INVENTION: IN A PROCESS FOR DETERMINATION OF ANTIBODIES OR BIOTINYLATED

; TITLE OF INVENTION: PEPTIDES CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES,

; NUMBER OF SEQUENCES: 453

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:


```
; APPLICATION NUMBER: US/08/146.028
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 26 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: HCV
; FEATURE:
;   NAME/KEY: Modified-site
;   LOCATION: 1
; FEATURE:
;   NAME/KEY: Modified-site
;   LOCATION: 26
; US-08-146-028-85

Query Match      100.0%; Score 20; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKONIQLINT 20
Db 6 SGLVSLFTPGAKONIQLINT 25

RESULT 6
US-08-723-425A-85
; Sequence 85, Application US/08723425A
; Patent No. 6165730
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT
; TITLE OF INVENTION: PROCESS FOR THE DETERMINATION OF
; TITLE OF INVENTION: PEPTIDES CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT
; TITLE OF INVENTION: EPITOPES AND THEIR USE IN A PROCESS FOR DETERMINATION OF
; TITLE OF INVENTION: ANTIBODIES OR BIOTINYLATED PEPTIDES CORRESPONDING ...
; NUMBER OF SEQUENCES: 453
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE, P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
; CITY: Arlington
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/723,425A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 1487-13
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 26 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: HCV
; FEATURE:
;   NAME/KEY: Modified-site
;   LOCATION: 1
```

```
; FEATURE:
;   NAME/KEY: Modified-site
;   LOCATION: 26
; US-08-723-425A-85

Query Match      100.0%; Score 20; DB 3; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKONIQLINT 20
Db 6 SGLVSLFTPGAKONIQLINT 25

RESULT 7
US-09-112-206-85
; Sequence 85, Application US/09112206
; Patent No. 6210903
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: PROCESS FOR THE DETERMINATION OF PEPTIDES
; TITLE OF INVENTION: CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES AND THEIR
; TITLE OF INVENTION: IN A PROCESS FOR DETERMINATION OF ANTIBODIES OR BIOTINYLATED
; TITLE OF INVENTION: PEPTIDES CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES,
; TITLE OF INVENTION: PROCESS FOR PREPARING THEM AND COMPOSITIONS CONTAINING THEM,
; NUMBER OF SEQUENCES: 453
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/112,206
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/146,028
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 26 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: HCV
; FEATURE:
;   NAME/KEY: Modified-site
;   LOCATION: 1
; FEATURE:
;   NAME/KEY: Modified-site
;   LOCATION: 26
; US-09-112-206-85

Query Match      100.0%; Score 20; DB 3; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKONIQLINT 20
Db 6 SGLVSLFTPGAKONIQLINT 25

RESULT 8
US-08-146-028-83
; Sequence 83, Application US/08146028
; Patent No. 5891640
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: PROCESS FOR THE DETERMINATION OF PEPTIDES
; TITLE OF INVENTION: CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES AND THEIR
; TITLE OF INVENTION: IN A PROCESS FOR DETERMINATION OF ANTIBODIES OR BIOTINYLATED
; TITLE OF INVENTION: PEPTIDES CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES,
```

;; TITLE OF INVENTION: PROCESS FOR PREPARING THEM AND COMPOSITIONS CONTAINING THEM
;; NUMBER OF SEQUENCES: 453
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/146,028
;; INFORMATION FOR SEQ ID NO: 83:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 36 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; HYPOTHETICAL: NO
;; ORIGINAL SOURCE:
;; INDIVIDUAL ISOLATE: HCV
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 1
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 36
;; US-08-146-028-83

Query Match 100.0%; Score 20; DB 2; Length 36;
Best Local Similarity 100.0%; Pred. No. 1.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKNQIQLINT 20
Db 16 SGLVSLFTPGAKNQIQLINT 35

RESULT 9

;; Sequence 131, Application US/08146028
;; Patent No. 5891640
;; GENERAL INFORMATION:
;; APPLICANT:
;; TITLE OF INVENTION: PROCESS FOR THE DETERMINATION OF PEPTIDES
;; CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES AND THEIR
;; TITLE OF INVENTION: IN A PROCESS FOR DETERMINATION OF ANTIBODIES OR BIOTINYLATED
;; TITLE OF INVENTION: PEPTIDES CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES,
;; TITLE OF INVENTION: PROCESS FOR PREPARING THEM AND COMPOSITIONS CONTAINING THEM
;; NUMBER OF SEQUENCES: 453
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/146,028
;; INFORMATION FOR SEQ ID NO: 131:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 36 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; HYPOTHETICAL: NO
;; ORIGINAL SOURCE:
;; INDIVIDUAL ISOLATE: HCV
;; FEATURE:
;; NAME/KEY: Xaa is absent
;; LOCATION: 1
;; FEATURE:
;; NAME/KEY: Xaa is absent
;; LOCATION: 36
;; US-08-146-028-131

Query Match 100.0%; Score 20; DB 2; Length 36;
Best Local Similarity 100.0%; Pred. No. 1.8e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SGLVSLFTPGAKNQIQLINT 20
Db 16 SGLVSLFTPGAKNQIQLINT 35
RESULT 10
US-08-723-425A-83
;; Sequence 83, Application US/08723425A
;; Patent No. 6165730
;; GENERAL INFORMATION:
;; APPLICANT: DELEYS, ROBERT
;; TITLE OF INVENTION: PROCESS FOR THE DETERMINATION OF
;; TITLE OF INVENTION: PEPTIDES CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT
;; TITLE OF INVENTION: EPITOPES AND THEIR USE IN A PROCESS FOR DETERMINATION OF
;; TITLE OF INVENTION: ANTIBODIES OR BIOTINYLATED PEPTIDES CORRESPONDING ...
;; NUMBER OF SEQUENCES: 453
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: NIXON & VANDERHUYE, P.C.
;; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
;; CITY: Arlington
;; STATE: VA
;; COUNTRY: USA
;; ZIP: 22201
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/723,425A
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: SADOFF, B.J.
;; REGISTRATION NUMBER: 36,663
;; REFERENCE/DOCKET NUMBER: 1487-13
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 703-816-4000
;; TELEFAX: 703-816-4100
;; INFORMATION FOR SEQ ID NO: 83:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 36 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; HYPOTHETICAL: NO
;; ORIGINAL SOURCE:
;; INDIVIDUAL ISOLATE: HCV
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 1
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 36
;; US-08-723-425A-83

Query Match 100.0%; Score 20; DB 3; Length 36;
Best Local Similarity 100.0%; Pred. No. 1.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKNQIQLINT 20
Db 16 SGLVSLFTPGAKNQIQLINT 35

RESULT 11

US-08-723-425A-131
;; Sequence 131, Application US/08723425A
;; Patent No. 6165730
;; GENERAL INFORMATION:
;; APPLICANT: DELEYS, ROBERT

;; TITLE OF INVENTION: PROCESS FOR THE DETERMINATION OF
;; TITLE OF INVENTION: PEPTIDES CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT
;; TITLE OF INVENTION: EPITOPES AND THEIR USE IN A PROCESS FOR DETERMINATION OF
;; TITLE OF INVENTION: ANTIBODIES OR BIOTINYLATED PEPTIDES CORRESPONDING ...
;; NUMBER OF SEQUENCES: 453
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: NIXON & VANDERHUYE, P.C.
;; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
;; CITY: Arlington
;; STATE: VA
;; COUNTRY: USA
;; ZIP: 22201
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/723.425A
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: SADOFF, B.J.
;; REGISTRATION NUMBER: 36,663
;; REFERENCE/DOCKET NUMBER: 1487-13
;; TELEPHONE: 703-816-4000
;; TELEFAX: 703-816-4100
;; INFORMATION FOR SEQ ID NO: 131:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 36 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; HYPOTHETICAL: NO
;; ORIGINAL SOURCE:
;; INDIVIDUAL ISOLATE: HCV
;; FEATURE:
;; NAME/KEY: Xaa is absent
;; LOCATION: 1
;; FEATURE:
;; NAME/KEY: Xaa is absent
;; LOCATION: 36
;; US-08-723-425A-131

Query Match 100.0%; Score 20; DB 3; Length 36;
Best Local Similarity 100.0%; Pred. No. 1.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLINT 20
Db 16 SGLVSLFTPGAKQNIQLINT 35

RESULT 12
US-09-112-206-83
; Sequence 83, Application US/09112206
; Patent No. 6210903
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: PROCESS FOR THE DETERMINATION OF PEPTIDES
; TITLE OF INVENTION: CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES AND THEIR
; TITLE OF INVENTION: IN A PROCESS FOR DETERMINATION OF ANTIBODIES OR BIOTINYLATED
; TITLE OF INVENTION: PEPTIDES CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES,
; TITLE OF INVENTION: PROCESS FOR PREPARING THEM AND COMPOSITIONS CONTAINING THEM
; NUMBER OF SEQUENCES: 453
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/112.206

;; FILING DATE:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/146,028
;; FILING DATE:
;; INFORMATION FOR SEQ ID NO: 83:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 36 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; HYPOTHETICAL: NO
;; ORIGINAL SOURCE:
;; INDIVIDUAL ISOLATE: HCV
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 1
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 36
;; US-09-112-206-83

Query Match 100.0%; Score 20; DB 3; Length 36;
Best Local Similarity 100.0%; Pred. No. 1.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SGLVSLFTPGAKQNIQLINT 20
Db 16 SGLVSLFTPGAKQNIQLINT 35

RESULT 13
US-09-112-206-131
; Sequence 131, Application US/09112206
; Patent No. 6210903
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: PROCESS FOR THE DETERMINATION OF PEPTIDES
; TITLE OF INVENTION: CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES AND THEIR
; TITLE OF INVENTION: IN A PROCESS FOR DETERMINATION OF ANTIBODIES OR BIOTINYLATED
; TITLE OF INVENTION: PEPTIDES CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES,
; TITLE OF INVENTION: PROCESS FOR PREPARING THEM AND COMPOSITIONS CONTAINING THEM
; NUMBER OF SEQUENCES: 453
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/112.206
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/146,028
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 131:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 36 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: HCV
; FEATURE:
; NAME/KEY: Xaa is absent
; LOCATION: 1
; FEATURE:
; NAME/KEY: Xaa is absent
; LOCATION: 36
; US-09-112-206-131

Query Match 100.0%; Score 20; DB 3; Length 36;
Best Local Similarity 100.0%; Pred. No. 1.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: November 21, 2003, 21:15:13
Job time : 10.6 secs

;
ADDRESSEE: Chiron Corporation
;
STREET: 4560 Horton Street

C;Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 17-Nov-2000
 C;Accession: A44150
 B;Ching, W.M.; Wyckowski, C.; Beach, M.J.; Wang, H.; Davies, C.L.; Carl, M.; Bradley, D.
 Proc. Natl. Acad. Sci. U.S.A. 89, 3190-3194, 1992
 A;Title: Interaction of immune sera with synthetic peptides corresponding to the structure
 A;Reference number: A44150; MUID:92228749; PMID:1373489
 A;Accession: A44150
 A;Status: preliminary; not compared with conceptual translation
 A;Molecule type: genomic RNA
 A;Residues: 1-513 <CHI>
 C;Superfamily: hepatitis C virus genome polyprotein

Query Match 65.0%; Score 13; DB 2; Length 513;
 Best Local Similarity 100.0%; Pred. No. 5.3e-06;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 TPGAKONTQLINT 20
 Db 404 TPGAKONTQLINT 416
 |||||

RESULT 3
 GNVVCH
 N;Contains: genome polyprotein - hepatitis C virus (strain H)
 N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C;Species: hepatitis C virus
 A;Note: host Homo sapiens (man)
 C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C;Accession: A36814; A41546
 R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 submitted to GenBank, July 1992
 A;Description: Genomic structure of the human prototype strain H of hepatitis C virus: C
 A;Reference number: A36814
 A;Accession: A36814
 A;Molecule type: genomic RNA
 A;Residues: 1-3011 <INC>
 A;Cross-references: GB:M67463; NID:q329737; PIDN:AAA45534.1; PID:q329738
 R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
 A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari
 A;Reference number: A41546; MUID:92052256; PMID:1658800
 A;Contents: annotation
 A;Note: neither amino acid nor nucleotide sequence is given
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
 F;1-115/Product: capsid protein C #status predicted <CPC>
 F;192-191/Product: envelope protein M #status predicted <EPM>
 F;192-389/Product: major envelope protein E #status predicted <MEE>
 F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F;1007-1615/Product: hepacivirin #status predicted <NS3>
 F;1230-1237/Region: nucleotide-binding motif A (P-loop)
 F;1312-1317/Region: nucleotide-binding motif B
 F;1316-1319/Region: DEXH motif
 F;1616-1862/Product: nonstructural protein NS4a #status predicted <NS4A>
 F;1863-2013/Product: nonstructural protein NS4b #status predicted <NS4B>
 F;2034-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match 65.0%; Score 13; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 2.4e-05;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 TPGAKONTQLINT 20
 Db 404 TPGAKONTQLINT 416
 |||||

RESULT 4
 D48776
 N;Contains: genome polyprotein - hepatitis C virus (fragment)
 C;Species: hepatitis C virus

C;Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 17-Nov-2000
 C;Accession: D48776
 R;Higashi, Y.; Kakumu, S.; Yoshioka, K.; Wakita, T.; Mizokami, M.; Ohba, K.; Ito, Y.; Is
 Virology 197, 659-668, 1993
 A;Title: Dynamics of genome change in the E2/NS1 region of hepatitis C virus in vivo.
 A;Reference number: A48776; MUID:94069940; PMID:8249288
 A;Accession: D48776
 A;Status: preliminary; not compared with conceptual translation
 A;Molecule type: nucleic acid
 A;Residues: 1-133 <HIG>
 A;Experimental source: subtype III, patient KS
 A;Note: sequence extracted from NCBI backbone (NCBIP:140217)
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: polyprotein

Query Match 60.0%; Score 12; DB 2; Length 133;
 Best Local Similarity 100.0%; Pred. No. 1.8e-05;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 PGAKONTQLINT 20
 Db 34 PGAKONTQLINT 45
 |||||

RESULT 5
 S24097
 N;Contains: envelope protein - hepatitis C virus (fragment)
 N;Contains: envelope protein gp35; envelope protein gp70
 C;Species: hepatitis C virus
 C;Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
 C;Accession: S24097; S78235
 R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraishi, K.; Ohkoshi,
 Virus Res. 22, 107-123, 1992
 A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C viru
 A;Reference number: S24068; MUID:92230394; PMID:1314471
 A;Accession: S24097
 A;Molecule type: genomic RNA
 A;Residues: 1-138 <KAT>
 A;Cross-references: EMBL:X60591
 A;Experimental source: isolate RE69
 R;Kato, N.
 submitted to the EMBL Data Library, August 1991
 A;Reference number: S78226
 A;Accession: S78235
 A;Molecule type: genomic RNA
 A;Residues: 1-107, 'E', 109-114, 'R', 116-138 <KAW>
 A;Cross-references: EMBL:X60591
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: envelope protein; glycoprotein; polyprotein
 F;1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
 F;16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
 F;16-42/Region: hypervariable 1 #status predicted
 F;106-112/Region: hypervariable 2 #status predicted
 F;49,55,62,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 45.0%; Score 9; DB 2; Length 138;
 Best Local Similarity 100.0%; Pred. No. 0.025;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLVSLFTPG 10
 Db 3Q GLVSLFTPG 38
 |||||

RESULT 6
 PS0449
 N;Contains: genome polyprotein - hepatitis C virus (strain PRC11) (fragments)
 N;Contains: envelope protein E2; nonstructural protein NS1; nonstructural protein NS5
 C;Species: hepatitis C virus
 C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
 C;Accession: PS0449
 R;Liu, K.; Hu, Z.; Li, H.; Prince, A.M.; Inchauspe, G.
 Gene 114, 245-250, 1992

RESULT 10
S24088
envelope protein - hepatitis C virus (fragment)
N;Contains: envelope protein gp35; envelope protein gp70
C;Species: hepatitis C virus
C;Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C;Accession: S24088; S78229
R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraio, K.; Ohkoshi
Virus Res. 22, 107-123, 1992
A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A;Reference number: S24068; MUID:52230394; PMID:1314471
A;Accession: S24088
A;Molecule type: genomic RNA
A;Residues: 1-137 <KAT>
A;Cross-references: EMBL:X60574
A;Experimental source: isolate RE55C
R;Kato, N.
submitted to the EMBL Data Library, August 1991
A;Reference number: S78226
A;Accession: S78229
A;Molecule type: genomic RNA
A;Residues: 1-15,'X',16-137 <RAW>

A;Cross-references: EMBL:X60574
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: envelope protein; glycoprotein; polyprotein
 F;1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
 F;16-137/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
 F;16-41/Region: hypervariable 1 #status predicted
 F;105-111/Region: hypervariable 2 #status predicted
 F;48,54,61,79/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 40.0%; Score 8; DB 2; Length 137;
 Best Local Similarity 100.0%; Pred. No. 0.27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 QNIQLINT 20
 |||||
 Db 40 QNIQLINT 47

RESULT 11
 S24100
 envelope protein - hepatitis C virus (fragment)
 N;Contains: envelope protein gp35; envelope protein gp70
 C;Species: hepatitis C virus
 C;Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
 C;Accession: S24100; S78234
 R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraisho, K.; Ohkoshi, Virus Res. 22, 107-123, 1992
 A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
 A;Reference number: S24068; MUID:92230394; PMID:1314471
 A;Accession: S24100
 A;Molecule type: genomic RNA
 A;Residues: 1-138 <KAT>
 A;Cross-references: EMBL:X60583
 A;Experimental source: isolate RE71
 R;Kato, N.
 submitted to the EMBL Data Library, August 1991
 A;Reference number: S78226
 A;Accession: S78234
 A;Molecule type: genomic RNA
 A;Residues: 1-83, C', 85-138 <KAW>
 A;Cross-references: EMBL:X60583
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: envelope protein; glycoprotein; polyprotein
 F;1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
 F;16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
 F;16-42/Region: hypervariable 1 #status predicted
 F;106-112/Region: hypervariable 2 #status predicted
 F;49,55,62,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 40.0%; Score 8; DB 2; Length 138;
 Best Local Similarity 100.0%; Pred. No. 0.28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 QNIQLINT 20
 |||||
 Db 41 QNIQLINT 48

RESULT 12
 S24096
 envelope protein - hepatitis C virus (fragment)
 N;Contains: envelope protein gp35; envelope protein gp70
 C;Species: hepatitis C virus
 C;Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
 C;Accession: S24096; S78233
 R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraisho, K.; Ohkoshi, Virus Res. 22, 107-123, 1992
 A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
 A;Reference number: S24068; MUID:92230394; PMID:1314471
 A;Accession: S24096
 A;Molecule type: genomic RNA
 A;Residues: 1-138 <KAT>
 A;Cross-references: EMBL:X60581

A;Experimental source: isolate RE68
 R;Kato, N.
 submitted to the EMBL Data Library, August 1991
 A;Reference number: S78226
 A;Accession: S78233
 A;Molecule type: genomic RNA
 A;Residues: 1-23, 'A', 25-40, 'Q', 42-67, 'G', 69-138 <KAW>
 A;Cross-references: EMBL:X60581
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: envelope protein; glycoprotein; polyprotein
 F;1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
 F;16-136/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
 F;16-42/Region: hypervariable 1 #status predicted
 F;106-112/Region: hypervariable 2 #status predicted
 F;49,55,62,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 40.0%; Score 8; DB 2; Length 138;
 Best Local Similarity 100.0%; Pred. No. 0.28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SLFTPGAK 12
 |||||
 Db 33 SLFTPGAK 40

RESULT 13
 PC1205
 envelope protein - hepatitis C virus (strain RS3-2) (fragment)
 C;Species: hepatitis C virus
 C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
 C;Accession: PC1205
 R;Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Sekiya, H.; Hijikata, M.; Shimotoh Biochem. Biophys. Res. Commun. 189, 119-127, 1992
 A;Title: Characterization of hypervariable regions in the putative envelope protein of h
 A;Reference number: PC1182; MUID:93080545; PMID:1333186
 A;Accession: PC1205
 A;Status: nucleic acid sequence not shown
 A;Molecule type: DNA
 A;Residues: 1-138 <KAT>
 A;Cross-references: DDBJ:D12942; DDBJ:D12972
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: envelope protein

Query Match 40.0%; Score 8; DB 2; Length 138;
 Best Local Similarity 100.0%; Pred. No. 0.28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 QNIQLINT 20
 |||||
 Db 41 QNIQLINT 48

RESULT 14
 PC1203
 envelope protein - hepatitis C virus (strain RS2-4) (fragment)
 C;Species: hepatitis C virus
 C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
 C;Accession: PC1203
 R;Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Sekiya, H.; Hijikata, M.; Shimotoh Biochem. Biophys. Res. Commun. 189, 119-127, 1992
 A;Title: Characterization of hypervariable regions in the putative envelope protein of h
 A;Reference number: PC1182; MUID:93080545; PMID:1333186
 A;Accession: PC1203
 A;Status: nucleic acid sequence not shown
 A;Molecule type: DNA
 A;Residues: 1-138 <KAT>
 A;Cross-references: DDBJ:D12942; DDBJ:D12972
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: envelope protein

Query Match 40.0%; Score 8; DB 2; Length 138;
 Best Local Similarity 100.0%; Pred. No. 0.28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 QNLIQINT 20
| | | | |
Db 41 QNLIQINT 48

RESULT 15

PC1206
envelope protein - hepatitis C virus (strain RS3-3) (fragment)
C:Species: hepatitis C virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C:Accession: PC1206
R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Sekiya, H.; Hijikata, M.; Shimotoh
Biochem. Biophys. Res. Commun. 189, 119-127, 1992
A:Title: Characterization of hypervariable regions in the putative envelope protein of h
A:Reference number: PC1182; MUID:93080545; PMID:1333186
A:Accession: PC1206
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-138 <KAT>
A:Cross-references: DDBJ:D12942; DDBJ:D12972
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein

Query Match 40.0%; Score 8; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 QNLIQINT 20
| | | | |
Db 41 QNLIQINT 48

Search completed: November 21, 2003, 21:11:28
Job time : 9.25 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:52 ; Search time 4.85 Seconds
(without alignments)
193.925 Million cell updates/sec

Title: US-09-973-025-72

Perfect score: 20

Sequence: 1 SGLVSLFTPGAKNIQLINT 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13	65.0	3011	1 POLG_HCVH	P27958 h genome po
2	11	55.0	138	1 POLG_HCVH	P27953 hepatitis c
3	8	40.0	321	1 POLG_HCVH	P27957 hepatitis c
4	7	35.0	298	1 SERB_SCHPO	O74382 schizosacch
5	7	35.0	2054	1 YCF2_PINTH	P41653 pinus thunb
6	6	30.0	66	1 CXA2_CONTE	O9xzk7 conus texti
7	6	30.0	69	1 CXA1_CONTE	O9xzk6 conus texti
8	6	30.0	86	1 CYSZ_PASHA	P72205 pasteurella
9	6	30.0	138	1 Y076_MYCGE	P47322 mycoplasma
10	6	30.0	146	1 YN59_YEAST	P53826 saccharomyc
11	6	30.0	166	1 RL21_ENTHI	P38653 entamoeba h
12	6	30.0	224	1 TCA2_VIBCH	P23024 vibrio chol
13	6	30.0	238	1 Y206_CHLMU	O9plal chlamydia m
14	6	30.0	238	1 Y819_CHLTR	O84826 chlamydia t
15	6	30.0	244	1 MTX2_CABEL	P34599 caenorhabdi
16	6	30.0	286	1 KSGA_STRCO	O9k3r5 s dimethyla
17	6	30.0	309	1 POLG_HCVH7	P27955 hepatitis c
18	6	30.0	310	1 LDH_THEAQ	P13715 thermus aqu
19	6	30.0	311	1 HA58_HALS8	O9hna8 haloarchaeo
20	6	30.0	317	1 MOCA_RHIME	P49307 rhizobium m
21	6	30.0	353	1 ML1A_PHOSU	P49217 phodopus su
22	6	30.0	373	1 DMT1_HUMAN	O9y5r6 homo sapien
23	6	30.0	373	1 DMT1_PIG	O9t01 sus scrofa
24	6	30.0	374	1 DMT1_MOUSE	O9qz59 mus musculu
25	6	30.0	381	1 YF03_MYCPN	P15283 mycoplasma
26	6	30.0	392	1 SERA_MANSE	P14754 manduca sex
27	6	30.0	436	1 SH6_RAT	P31388 rattus norv
28	6	30.0	440	1 SH6_HUMAN	P50406 homo sapien
29	6	30.0	440	1 SH6_MOUSE	O9rlc8 mus musculu
30	6	30.0	493	1 YE14_MYCPN	P75372 mycoplasma
31	6	30.0	513	1 POLG_HCVJ2	P27959 hepatitis c
32	6	30.0	520	1 POLG_HCVH4	O01404 hepatitis c
33	6	30.0	520	1 POLG_HCVHK	Q01403 hepatitis c

34 6 30.0 526 1 PUR9_BUCAP Q8ka70 b bifunctio
35 6 30.0 549 1 YJCE_ECOLI P32703 escherichia
36 6 30.0 552 1 PYRG_WIGBR Q8d2k0 wigglewort
37 6 30.0 582 1 NIFA_BRAJA P05407 bradyrhizob
38 6 30.0 582 1 SHO2_HUMAN Q8uq13 homo sapien
39 6 30.0 582 1 SHO2_MOUSE O88520 mus musculu
40 6 30.0 585 1 YB13_YEAST P38289 saccharomyc
41 6 30.0 597 1 BP10_PARLI P42674 paracentrot
42 6 30.0 659 1 DNLJ_MYCGE P47496 mycoplasma
43 6 30.0 710 1 TGM7_HUMAN O96pf1 homo sapien
44 6 30.0 762 1 KGP2_HUMAN Q13237 homo sapien
45 6 30.0 762 1 KGP2_MOUSE Q61410 mus musculu

ALIGNMENTS

RESULT 1

POLG_HCVH
ID POLG_HCVH STANDARD; PRT; 3011 AA.
AC P27958;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate H) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658800;
RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C
RT virus: comparison with American and Japanese isolates.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX MEDLINE=97331322; PubMed=9187654;
RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain.";
RL Nat. Struct. Biol. 4:463-467(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE=98154321; PubMed=9493270;
RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding";
RL Structure 6:89-100(1998).
CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1

CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M67463; AAA45534.1; --
 DR PIR; A36814; GNVVCH.
 DR PDB; 1HEI; 25-NOV-98.
 DR PDB; 1A1V; 16-FEB-99.
 DR PDB; 1A1R; 17-JUN-98.
 DR MEROPS; S29.001; --
 DR MEROPS; U39.001; --
 DR TRANSFAC; T04155; --
 DR TRANSFAC; IPRO01410; DEAD.
 DR InterPro; IPRO02522; HCV capsid.
 DR InterPro; IPRO02521; HCV core.
 DR InterPro; IPRO02519; HCV env.
 DR InterPro; IPRO02531; HCV NS1.
 DR InterPro; IPRO02518; HCV NS2.
 DR InterPro; IPRO04109; HCV NS3.
 DR InterPro; IPRO00745; HCV NS4a.
 DR InterPro; IPRO01490; HCV NS4b.
 DR InterPro; IPRO02868; HCV NS5a.
 DR InterPro; IPRO02166; HCV RdRP.
 DR InterPro; IPRO01650; Helicase_C.
 DR InterPro; IPRO07095; RNA pol_DS ps.
 DR InterPro; IPRO07094; RNA pol_Psvir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 192 283 CAPSID PROTEIN C.
 FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
 FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
 FT CHAIN 810 1026 PROTEIN P7.
 FT CHAIN 1027 1657 PROTEASE/HELICASE NS3.
 FT CHAIN 1658 1711 NONSTRUCTURAL PROTEIN NS2.
 FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4A.
 FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4B.
 FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5A.
 FT TRANSMEM 347 369 NONSTRUCTURAL PROTEIN NS5B.
 FT ACT_SITE 1083 1083 POTENTIAL.
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1224 1226
 FT TURN 1232 1233
 FT TURN 1236 1238
 FT HELIX 1239 1246
 FT TURN 1247 1248
 FT STRAND 1251 1255
 FT HELIX 1258 1271
 FT TURN 1272 1272
 FT STRAND 1277 1280
 FT TURN 1281 1282
 FT STRAND 1283 1285
 FT STRAND 1291 1295
 FT HELIX 1296 1301
 FT TURN 1302 1303
 FT STRAND 1312 1316
 FT TURN 1317 1319
 FT HELIX 1323 1335
 FT TURN 1336 1340
 FT STRAND 1343 1347
 FT TURN 1352 1353
 FT TURN 1360 1361
 FT STRAND 1362 1366
 FT STRAND 1368 1368
 FT STRAND 1373 1375
 FT TURN 1376 1377
 FT STRAND 1378 1380
 FT HELIX 1382 1385
 FT STRAND 1389 1393
 FT HELIX 1397 1409
 FT TURN 1410 1411
 FT STRAND 1414 1417
 FT TURN 1419 1420
 FT STRAND 1432 1436
 FT TURN 1438 1439
 FT STRAND 1450 1453
 FT STRAND 1456 1463
 FT STRAND 1471 1478
 FT STRAND 1480 1480
 FT HELIX 1481 1488
 FT TURN 1489 1490
 FT STRAND 1497 1501
 FT STRAND 1507 1507
 FT STRAND 1511 1511
 FT HELIX 1514 1527
 FT STRAND 1532 1544
 FT STRAND 1550 1550
 FT HELIX 1555 1564
 FT HELIX 1570 1578
 FT TURN 1579 1580
 FT HELIX 1584 1597
 FT TURN 1598 1598
 FT HELIX 1606 1611
 FT TURN 1614 1618
 FT STRAND 1622 1623
 FT STRAND 1627 1627
 FT STRAND 1635 1636
 FT HELIX 1640 1652
 SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;
 Query Match 65.0%; Score 13; DB 1; Length 3011;

RESULT 3
POLG HCVTH

15-SEP-2003 (Rel. 42, Last annotation update)
Probable phosphoserine phosphatase (EC 3.1.3.3) (PSP) (O-phosphoserine phosphohydrolase) (PSPase).

OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]

SEQUENCE FROM N.A.

RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A., Brooks J., Peat N., Hayes J., Baker S., Basham D., Bowman S., Collins M., Brown D., Brown S., Chillingworth T., Churcher C.M., Gentles S., Connor R., Cronin A., Davis P., Feltwell T., Fraser A., Gattis S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K., James K., Jones L., Jones M., Leather S., McDonald S., McLean J., Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C., Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E., Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S., Skellon J., Simmonds M., Squares R., Squares S., Stevens K., Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B., Weljens J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H., Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B., Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G., Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L., Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J., Shpakovski G.V., Useery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomycetes pombe";
RL Nature 415:871-880(2002).
RN [2]

SEQUENCE FROM N.A.

RC STRAIN=PR745;
RX MEDLINE=98162722; PubMed=9501991;
RA Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
RT "Identification of open reading frames in Schizosaccharomycetes pombe CDNA8";
RL DNA Res. 4:363-369(1997).
CC -!- CATALYTIC ACTIVITY: Phosphoserine + H(2)O = serine + phosphate.
CC -!- SIMILARITY: BELONGS TO THE SERB FAMILY.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a frameshift in position 267.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AL031261; CAA20303.1; --
EMBL; D89261; BAA13922.1; ALT_FRAME.
PIR; T40408; T40408
GeneDB_Spombe; SPBC3H7.07c; --
InterPro; IPR006383; HAD_SF_IB.
InterPro; IPR005834; Hydrolase.
InterPro; IPR004469; SerB.
Pfam; PF00702; Hydrolase; 1.
TIGRFAMs; TIGR01488; HAD-SF-IB; 1.
TIGRFAMs; TIGR00338; SerB; 1.
KW Hydrolase; Serine biosynthesis.
SEQUENCE 298 AA; 32411 MW; 48CBCE73BE9013CA5 CRC64;

Query Match 35.0%; Score 7; DB 1; Length 298;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 FTPGAQK 13
| | | | |
Db 148 FTPGAQK 154

RESULT 5

YCF2_PINTH STANDARD; PRT; 2054 AA.
ID AC P41653;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 244.6 kDa protein ycf2 (ORF 2054).
GN YCF2.
OS Pinus thunbergii (Green pine) (Japanese black pine).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=3350;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95024047; PubMed=7937893;
RA Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T., Sugitara M.;
RT "Loss of all ndh genes as determined by sequencing the entire chloroplast genome of the black pine Pinus thunbergii";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
CC -!- FUNCTION: NOT YET KNOWN.
CC -!- SIMILARITY: BELONGS TO THE YCF2 FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; D17510; BAA04460.1; --
PIR; T07584; T07584.
InterPro; IPR003593; AAA_ATPase.
InterPro; IPR003959; AAA_ATPase_centre.
Pfam; PF00004; AAA; 1.
SMART; SM00382; AAA; 1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2054 AA; 244604 MW; 6F5E92D078E33A9A CRC64;

Query Match 35.0%; Score 7; DB 1; Length 2054;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 FTPGAQK 13
| | | | |
Db 1767 FTPGAQK 1773

RESULT 6

CXA2_CONTE STANDARD; PRT; 66 AA.
ID AC Q9XZK7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alpha-type conotoxin Tx2 precursor.
OS Conus textile (Cloth-of-gold cone).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Orthogastropoda;
OC Apogastropoda; Caenogastropoda; Sorbeoconcha; Hypogastropoda;
OC Neogastropoda; Conoidea; Conidae; Conus.
OX NCBI_TaxID=6494;

RN SEQUENCE FROM N.A.
 RP TISSUE=Venom duct;
 RC MEDLINE=20037955; PubMed=10573284;
 RA Lu B.-S., Yu F., Zhao D., Huang P.-T., Huang C.-F.;
 RX "Conopeptides from *Conus striatus* and *Conus textile* by cDNA
 RT cloning."
 RL Peptides 20:1139-1144(1999).
 CC -!- FUNCTION: Alpha-conotoxins act on postsynaptic membranes, they
 CC bind to the nicotinic acetylcholine receptors (nAChR) and thus
 CC inhibit them (By similarity).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Expressed by the venom duct.
 CC -!- SIMILARITY: BELONGS TO THE A-SUPERFAMILY OF CONOTOXINS. ALPHA-TYPE
 CC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF146353; AAD31913.1; -.
 KW Postsynaptic neurotoxin; Neurotoxin; Toxin;
 KW Acetylcholine receptor inhibitor; Signal.
 FT SIGNAL 1 21 POTENTIAL.
 FT PROPEP 22 49 BY SIMILARITY.
 FT PEPTIDE 50 66 ALPHA-TYPE CONOTOXIN TX2.
 FT DISULFID 51 57 BY SIMILARITY.
 FT DISULFID 52 65 BY SIMILARITY.
 SQ SEQUENCE 66 AA; 7254 MW; EDD859BBAB94F26F CRC64;
 Query Match 30.0%; Score 6; DB 1; Length 66;
 Best Local Similarity 100.0%; Pred. No. 7.4;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SGLVSL 6
 DB 39 SGLVSL 44
 RESULT 7
 ID CXAL CONTE STANDARD; PRT; 69 AA.
 AC Q9VZK6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Alpha-type conotoxin Tx1 precursor.
 OS *Conus textile* (Cloth-of-gold cone).
 CC Eukaryota; Metazoa; Mollusca; Gastropoda; Orthogastropoda;
 CC Apogastropoda; Caenogastropoda; Sorbeoconcha; Hypogastropoda;
 CC Neogastropoda; Conoidea; Conidae; Conus.
 OX NCBI_TaxID=6494;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Venom duct;
 RA MEDLINE=20037955; PubMed=10573284;
 RX Lu B.-S., Yu F., Zhao D., Huang P.-T., Huang C.-F.;
 RT "Conopeptides from *Conus striatus* and *Conus textile* by cDNA
 RT cloning."
 RL Peptides 20:1139-1144(1999).
 CC -!- FUNCTION: Alpha-conotoxins act on postsynaptic membranes, they
 CC bind to the nicotinic acetylcholine receptors (nAChR) and thus
 CC inhibit them (By similarity).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Expressed by the venom duct.
 CC -!- SIMILARITY: BELONGS TO THE A-SUPERFAMILY OF CONOTOXINS. ALPHA-TYPE
 CC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF146352; AAD31912.1; -.
 KW Postsynaptic neurotoxin; Neurotoxin; Toxin;
 KW Acetylcholine receptor inhibitor; Signal; Amidation.
 FT SIGNAL 1 21 POTENTIAL.
 FT PROPEP 22 49 BY SIMILARITY.
 FT PEPTIDE 50 66 ALPHA-TYPE CONOTOXIN TX1.
 FT DISULFID 51 57 BY SIMILARITY.
 FT DISULFID 52 65 BY SIMILARITY.
 FT MOD RES 66 66 AMIDATION (G-67 PROVIDE AMIDE GROUP).
 SQ SEQUENCE 69 AA; 7442 MW; E36ECE90BF1B5580 CRC64;
 Query Match 30.0%; Score 6; DB 1; Length 69;
 Best Local Similarity 100.0%; Pred. No. 7.7;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SGLVSL 6
 DB 39 SGLVSL 44
 RESULT 8
 ID CYSZ_PASHA STANDARD; PRT; 86 AA.
 AC P72205;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE CysZ protein homolog (Fragment).
 GN CysZ.
 OS *Pasteurella haemolytica*.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 CC Pasteurellaceae; Mannheimia.
 OX NCBI_TaxID=75985;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Serotype A1 / ATCC 43270;
 RA Graham M.R., Lo R.Y.C.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: POSSIBLY INVOLVED IN SULFATE TRANSPORT.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Potential).
 CC -!- SIMILARITY: BELONGS TO THE CYSZ FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U62565; AAB09531.1; -.
 DR HAMAP; MF_00468; -; 1.
 DR Pfam; PF04401; DUF540; 1.
 KW Cysteine biosynthesis; Transmembrane; Inner membrane; Transport.
 FT NON TER 1
 FT TRANSMEM 7 27 POTENTIAL.
 SQ SEQUENCE 86 AA; 9602 MW; 6EBFAD1E9AE7F7FC CRC64;
 Query Match 30.0%; Score 6; DB 1; Length 86;
 Best Local Similarity 100.0%; Pred. No. 9.3;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 LVSLFT 8
 DB 8 LVSLFT 13

```

RESULT 9
Y076 MYCGE
ID Y076 MYCGE STANDARD; PRT; 138 AA.
AC P47322;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG076.
GN MG076.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Karhage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Doughterty B.A., Bock K.F., Hu P.-C., Lucier T.S.,
RA Tomb J.F., Dougherty B.A., Bock K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U36688; AAC1294.1; -.
DR PIR; D64208; D64208.
DR TIGR; MG076; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 17 37 POTENTIAL.
FT TRANSMEM 43 63 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
SQ SEQUENCE 138 AA; 15867 MW; 204E568264CCB8C0 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 138;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGLVSL 6
DB 132 SGLVSL 137

RESULT 10
YN59 YEAST
ID YN59 YEAST STANDARD; PRT; 146 AA.
AC P53826;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 16.1 kDa protein in KRE1-HXT14 intergenic region
DE precursor.
GN YNL319W OR N0343.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1676;
RX MEDLINE=96076632; PubMed=7502593;
RA Maftahi M., Nicaud J.-M., Levesque H., Gaillardin C.;

```

```

RT "Sequencing analysis of a 24.7 kb fragment of yeast chromosome XIV
RT identifies six known genes, a new member of the hexose transporter
RT family and ten new open reading frames.";
RL Yeast 11:1077-1085(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z46259; -. NOT ANNOTATED_CDS.
DR EMBL; Z71595; CAA96249.1; -.
DR PIR; S59575; S59575.
DR SGD; S0005263; YNL319W.
KW Hypothetical protein; Signal.
FT CHAIN 1 24 POTENTIAL.
FT SIGNAL 25 146 HYPOTHETICAL PROTEIN YNL319W.
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 106 106 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 146 AA; 16051 MW; FA232F00B97551B9 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 NIQLIN 19
DB 80 NIQLIN 85

RESULT 11
RL21 ENTHI
ID RL21 ENTHI STANDARD; PRT; 166 AA.
AC P38653;
DT 01-FEB-1996 (Rel. 31, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE 60S ribosomal protein L21.
GN RPL21.
OS Entamoeba histolytica.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=5759;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HM-1:IMSS, and 200:NIH;
RX MEDLINE=93133221; PubMed=1484554;
RA Petter R., Rozenblatt S., Nuchamowitz Y., Mirelman D.;
RT "Linkage between actin and ribosomal protein L21 genes in Entamoeba
RT histolytica.";
RL Mol. Biochem. Parasitol. 56:329-333(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HM-1:IMSS;
RX MEDLINE=95047535; PubMed=7959049;
RA Petter R., Moshitch S., Rozenblatt S., Nuchamowitz Y., Mirelman D.;
RT "Characterization of two distinct gene transcripts for ribosomal
RT protein L21 from pathogenic and nonpathogenic strains of Entamoeba
RT histolytica.";
RL Gene 150:181-186(1994).
CC -!- SIMILARITY: BELONGS TO THE L21E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M92098; AAA29116.1; -.

```



```

DR EMBL; X73634; CAA52014.1; -.
DR PIR; A48465; A48465.
DR InterPro; IPR001147; Ribosomal_L21e.
DR Pfam; PF01157; Ribosomal_L21e.1.
DR PROSITE; PS01171; RIBOSOMAL_L21E; 1.
KW Ribosomal protein.
FT VARIANT 113 113 A -> D (IN STRAIN 200:NIH).
FT VARIAT 116 116 R -> T (IN STRAIN 200:NIH).
FT CONFLICT 97 97 R -> K (IN REF. 1).
SQ SEQUENCE 166 AA; 18857 MW; FFA46E2335F9E5421 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 166;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 KQNIQL 17
Db 117 KQNIQL 122

RESULT 12
TCA2 VIBCH
ID TCA2 VIBCH STANDARD; PRT; 224 AA.
AC P23024;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Toxin co-regulated pilin precursor (Pilus colonization factor).
GN TCPA.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Classical Inaba 217561 / Serotype O1;
RX MEDLINE=90152368; PubMed=2576015;
RA Faast R., Ogierman M.A., Strother U.H., Manning P.A.;
RT "Nucleotide sequence of the structural gene, tcpA, for a major pilin
RT subunit of Vibrio cholerae.";
RL Gene 85:227-231(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Classical Inaba 217561 / Serotype O1;
RX MEDLINE=96200848; PubMed=8621096;
RA Ogierman M.A., Voss E., Meaney C., Faast R., Attridge S.R.,
RA Manning P.A.;
RT "Comparison of the promoter proximal regions of the toxin-co-regulated
RT tcp gene cluster in classical and El Tor strains of Vibrio cholerae
RT O1.";
RL Gene 170:9-16(1996).
RN [3]
RP SEQUENCE OF 26-50.
RX MEDLINE=87204222; PubMed=2883655;
RA Taylor R.K., Miller V.L., Furlong D.B., Mekalanos J.J.;
RT "Use of phoA gene fusions to identify a pilus colonization factor
RT coordinately regulated with cholera toxin.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:2833-2837(1987).
CC -!- FUNCTION: CONSTITUENT OF PILI, WHICH MAY BE INVOLVED IN ADHESION
CC OF V. CHOLERA TO THE HOST INTESTINAL EPITHELIUM.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: WEAK, TO THE N-ME-PHE PILI FAMILY. THE LEADER SEQUENCE
CC REGION AND SOME OTHER SEQUENCE PARTICULARITIES SUGGEST THAT TCPA
CC MAY REPRESENT A NOVEL CLASS OF PILIN, AND IMPLY THE EXISTENCE OF A
CC NOVEL SIGNAL PEPTIDASE.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----

```

```

CC EMBL; M33514; AAA86888.1; -.
DR EMBL; X64098; CAA45455.1; -.
DR PDB; 1QOZ; 14-JUN-02.
DR InterPro; IPR001120; Prok_N_methyltn.
DR PROSITE; PS00409; PROKAR_NTER_METHYL; 1.
KW Fimbria; Methylation; 3D-structure.
FT PROPEP 1 25 ATYPICAL LEADER PEPTIDE.
FT CHAIN 26 224 TOXIN CO-REGULATED PILIN.
FT MOD RES 26 26 METHYLATION (POTENTIAL).
FT DISULFID 145 211 BY SIMILARITY.
FT CONFLICT 57 57 I -> N (IN REF. 2).
FT CONFLICT 114 114 N -> I (IN REF. 2).
SQ SEQUENCE 224 AA; 23328 MW; E7468D3E27276B5 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 224;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SGLVSL 6
Db 96 SGLVSL 101

RESULT 13
Y206 CHLMU
ID Y206 CHLMU STANDARD; PRT; 238 AA.
AC Q9PLA1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TC0206.
GN TC0206.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Nigg;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linth K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AK39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE B11 FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; AB002287; AAF39078.1; -.
DR PIR; E81729; E81729.
DR TIGR; TC0206; -.
DR InterPro; IPR006214; UPF0005.
DR Pfam; PF01027; UPF0005; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 22 42 POTENTIAL.
FT TRANSMEM 49 69 POTENTIAL.
FT TRANSMEM 78 98 POTENTIAL.
FT TRANSMEM 105 125 POTENTIAL.
FT TRANSMEM 141 161 POTENTIAL.
FT TRANSMEM 166 186 POTENTIAL.
FT TRANSMEM 208 228 POTENTIAL.
SQ SEQUENCE 238 AA; 26309 MW; 4625B1D88093657E CRC64;

```

```

Query Match      30.0%; Score 6; DB 1; Length 238;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 VSLFTP 9
Db      159 VSLFTP 164

RESULT 14
Y819 CHLTR
ID Y819 CHLTR STANDARD; PRT; 238 AA.
AC 084826;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein CT819.
GN CT819.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/Cx;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE BIL FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001354; AAC68416.1; -.
DR PIR; E71467; E71467. UPF0005.
DR InterPro; IPR006214; UPF0005; 1.
DR Pfam; PF01027; UPF0005; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 22 42 POTENTIAL.
FT TRANSMEM 49 69 POTENTIAL.
FT TRANSMEM 78 98 POTENTIAL.
FT TRANSMEM 105 125 POTENTIAL.
FT TRANSMEM 141 161 POTENTIAL.
FT TRANSMEM 166 186 POTENTIAL.
FT TRANSMEM 208 228 POTENTIAL.
SQ SEQUENCE 238 AA; 26264 MW; FD101F7B3867B8E5 CRC64;

Query Match      30.0%; Score 6; DB 1; Length 238;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 VSLFTP 9
Db      159 VSLFTP 164

RESULT 15
MTX2 CAEEL
ID MTX2 CAEEL STANDARD; PRT; 244 AA.
AC F34599;
DT 01-FEB-1994 (Rel. 28, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

```

```

DE Metaxin 2 homolog.
GN ZC97.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RN Nature 368:32-38(1994).
RP [2]
RP REVISIONS.
RA Waterston R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Mitochondrial outer membrane (Potential).
CC -!- SIMILARITY: BELONGS TO THE METAXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L14714; AAA28180.2; -.
DR WormPep; ZC97.1; CE25677.
DR InterPro; IPR004046; GST_Cterm.
KW Hypothetical protein; Mitochondrion; Outer membrane; Transport;
KW Protein transport.
SQ SEQUENCE 244 AA; 28161 MW; 7021C151E00C00FC CRC64;

Query Match      30.0%; Score 6; DB 1; Length 244;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 VSLFTP 9
Db      10 VSLFTP 15

Search completed: November 21, 2003, 20:59:51
Job time : 5.85 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:45:57 ; Search time 24.4 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-72

Perfect score: 20

Sequence: 1 SGLVSLFTPGAKQNIQLINT 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	513	12 Q81220	Q81220 hepatitis c
2	20	100.0	3011	12 Q03463	Q03463 hepatitis c
3	16	80.0	169	12 Q91785	Q91785 hepatitis c
4	16	80.0	169	12 Q91788	Q91788 hepatitis c
5	16	80.0	169	12 Q91789	Q91789 hepatitis c
6	16	80.0	170	12 Q8QR78	Q8qr78 hepatitis c
7	15	75.0	154	12 Q91789	Q917b9 hepatitis c
8	15	75.0	154	12 Q917B7	Q917b7 hepatitis c
9	15	75.0	154	12 Q8QQQ8	Q8qqq8 hepatitis c
10	15	75.0	154	12 Q917B5	Q917b5 hepatitis c
11	15	75.0	154	12 Q917D4	Q917d4 hepatitis c
12	15	75.0	154	12 Q917D7	Q917d7 hepatitis c
13	15	75.0	154	12 Q8QQQ3	Q8qqq3 hepatitis c
14	15	75.0	154	12 Q8QQQ4	Q8qqq4 hepatitis c
15	15	75.0	154	12 Q917D0	Q917d0 hepatitis c
16	15	75.0	154	12 Q917D2	Q917d2 hepatitis c

ALIGNMENTS

RESULT 1

Q81220	PRELIMINARY;	PRT;	513 AA.
ID			
AC Q81220;			
DT 01-NOV-1996 (TRENBLrel. 01, Created)			
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)			
DE 01-OCT-2002 (TRENBLrel. 22, Last annotation update)			
OC Structural protein (Genome polyprotein) (Fragment).			
OS Hepatitis C virus.			
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC Hepacivirus.			
OX NCBI_TaxID=11103;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=HC-J1;			
RX MEDLINE=91013116; PubMed=2170712;			
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,			
RA Yoshizawa H.;			
RT "The 5'-terminal sequence of the hepatitis C virus genome.";			
RL Jpn. J. Exp. Med. 60:167-177(1990).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=HC-J1;			
RX MEDLINE=9204440; PubMed=1658196;			
RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,			
RA Miyakawa Y., Mayumi M.;			
RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions.";			
RT J. Gen. Virol. 72:2697-2704(1991).			
RL [3]			
RN SEQUENCE FROM N.A.			
RC STRAIN=HC-J1;			
RX MEDLINE=94270990; PubMed=7545932;			
RA Hotta H., Doi H., Hayaishi T., Purwanta M., Soemarto W., Mizokami M.,			
RA Ohba K., Homma M.;			
RT "Analysis of the core and E1 envelope region sequences of a novel variant of hepatitis C virus obtained in Indonesia.";			
RT Arch. Virol. 136:53-62(1994).			

Q917c3 hepatitis c
Q917d1 hepatitis c
Q917c6 hepatitis c
Q917b6 hepatitis c
Q8qqq5 hepatitis c
Q917d5 hepatitis c
Q917c4 hepatitis c
Q8qqg6 hepatitis c
Q917c1 hepatitis c
Q917b8 hepatitis c
Q917b4 hepatitis c
Q917c9 hepatitis c
Q917d6 hepatitis c
Q8bbp3 hepatitis c
Q8bb5 hepatitis c
Q8bb4 hepatitis c
Q8bb2 hepatitis c
Q8bb1 hepatitis c
Q8bb0 hepatitis c
Q8bbq9 hepatitis c
Q8bbq8 hepatitis c
Q8bbq7 hepatitis c
Q8bbq6 hepatitis c
Q8bbq5 hepatitis c
Q8bbq2 hepatitis c
Q8bbp5 hepatitis c
Q8bbp1 hepatitis c
Q8bb14 hepatitis c

17 15 75.0 154 12 Q917C3
18 15 75.0 154 12 Q917D1
19 15 75.0 154 12 Q917C6
20 15 75.0 154 12 Q917B6
21 15 75.0 154 12 Q8QQQ7
22 15 75.0 154 12 Q8QQQ5
23 15 75.0 154 12 Q917D5
24 15 75.0 154 12 Q917C4
25 15 75.0 154 12 Q8QQQ6
26 15 75.0 154 12 Q917C1
27 15 75.0 154 12 Q917B8
28 15 75.0 154 12 Q917B4
29 15 75.0 154 12 Q917C9
30 15 75.0 154 12 Q917D6
31 15 75.0 187 12 Q8BBP3
32 15 75.0 191 12 Q8BBR5
33 15 75.0 191 12 Q8BBR4
34 15 75.0 191 12 Q8BBR2
35 15 75.0 191 12 Q8BBR1
36 15 75.0 191 12 Q8BBR0
37 15 75.0 191 12 Q8BBQ9
38 15 75.0 191 12 Q8BBQ8
39 15 75.0 191 12 Q8BBQ7
40 15 75.0 191 12 Q8BBQ6
41 15 75.0 191 12 Q8BBQ5
42 15 75.0 191 12 Q8BBQ2
43 15 75.0 191 12 Q8BBP5
44 15 75.0 191 12 Q8BBP1
45 15 75.0 191 12 Q8BB14

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).

DR EMBL; D00831; BAA00705.1; -.
DR InterPro; IPR002522; HCV_capsid.

DR InterPro; IPR002521; HCV core.

DR InterPro; IPR002519; HCV env.

DR InterPro; IPR002531; HCV NS1.

DR Pfam; PF01543; HCV_capsid; 1.

DR Pfam; PF01542; HCV_core; 1.

DR Pfam; PF01539; HCV env; 1.

DR Pfam; PF01560; HCV NS1; 1.

DR ProDom; PD186062; HCV NS1; 1.

KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;

KW Polyprotein; Transmembrane.

FT NON_TER 513 513

SEQUENCE 513 AA; 55710 MW; B1B54EBF9C359407 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 513;

Best Local Similarity 100.0%; Pred. No. 4e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGLVSLFTPGAKONIQLINT 20

Db 397 SGLVSLFTPGAKONIQLINT 416

RESULT 2

Q03463

ID Q03463 PRELIMINARY; PRT; 3011 AA.

AC Q03463;

DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE Genome polyprotein.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=111103;

RN [1]_TaxID=111103;

RP SEQUENCE FROM N.A.

RC STRAIN=HC-J1;

RX MEDLINE=91013116; PubMed=2170712;

RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,

RA Yoshizawa H.;

RT "The 5'-terminal sequence of the hepatitis C virus genome.";

RL Jpn. J. Exp. Med. 60:167-177(1990).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=HC-J1;

RX MEDLINE=92044440; PubMed=1658196;

RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,

RA Miyakawa Y., Mayumi M.;

RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated

RT from a human carrier: comparison with reported isolates for conserved

RT and divergent regions.";

RL J. Gen. Virol. 72:2697-2704(1991).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=HC-J1;

RX MEDLINE=93117120; PubMed=1335573;

RA Okamoto H., Kanai N., Mishihiro S.;

RT "Full-length nucleotide sequence of a Japanese hepatitis C virus

RT isolate (HC-J1) with high homology to USA isolates.";

RL Nucleic Acids Res. 20:6410-6410(1992).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=HC-J1;

RA Okamoto H.;

RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RC STRAIN=HC-J1;

RX MEDLINE=94174722; PubMed=7510436;

RA Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,

RA Inchauspe G.;

RT "Characterization and mapping of a B-cell immunogenic domain in

RT hepatitis C virus E2 glycoprotein using a yeast peptide library.";

RL Virology 200:246-255(1994).

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND MRNA (BY SIMILARITY).

DR EMBL; D10749; BAA01582.1; -.
DR HSSP; P27958; 1HEI.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV capsid.

DR InterPro; IPR002521; HCV_core.

DR InterPro; IPR002519; HCV env.

DR InterPro; IPR002531; HCV NS1.

DR InterPro; IPR002518; HCV NS2.

DR InterPro; IPR004109; HCV NS3.

DR InterPro; IPR000745; HCV NS4a.

DR InterPro; IPR003490; HCV NS4b.

DR InterPro; IPR002868; HCV NS5a.

DR InterPro; IPR002166; HCV RdRP.

DR InterPro; IPR001650; Helicase C.

DR InterPro; IPR007095; RNA_pol_DS_PS.

DR InterPro; IPR007094; RNA_pol_PSVir.

DR Pfam; PF01543; HCV_capsid; 1.

DR Pfam; PF01542; HCV_core; 1.

DR Pfam; PF01539; HCV env; 1.

DR Pfam; PF01560; HCV NS1; 1.

DR Pfam; PF01538; HCV NS2; 1.

DR Pfam; PF02907; HCV NS3; 1.

DR Pfam; PF02907; HCV NS3; 1.

DR Pfam; PF01006; HCV NS4a; 1.

DR Pfam; PF01001; HCV NS4b; 1.

DR Pfam; PF01506; HCV NS5a; 1.

DR Pfam; PF00271; Helicase C; 1.

DR Pfam; PF00998; Viral RdRP; 1.

DR ProDom; PD186062; HCV NS1; 1.

DR SMART; SM00487; DEXDc_1.

DR PROSITE; PS05057; RDRP_POSITIVE; 1.

DR PROSITE; PS05021; RDRP_VIRAL; 1.

KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;

KW Hydrolase; Nonstructural protein; Polyprotein;

KW RNA-directed RNA polymerase; Transferase; Transmembrane.

SQ SEQUENCE 3011 AA; 327112 MW; 97E9052C0250463B CRC64;

Query Match 100.0%; Score 20; DB 12; Length 3011;

Best Local Similarity 100.0%; Pred. No. 1.9e-12;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SGLVSLFTPGAKONIQLINT 20

Db 397 SGLVSLFTPGAKONIQLINT 416

RESULT 3

Q917R5

ID Q917R5 PRELIMINARY; PRT; 169 AA.

AC Q917R5;

DT 01-DEC-2001 (TReMBLrel. 19, Created)

DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)

DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)

DE Genome polyprotein (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=111103;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Hv1439-5;

RA Feucht H.;

RT "Long-term evolution of the hepatitis C virus hypervariable region.";

```
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF416529; AAL10425.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 169
SQ SEQUENCE 169 AA; 18331 MW; 055C30D086223696A CRC64;

Query Match 80.0%; Score 16; DB 12; Length 169;
Best Local Similarity 100.0%; Pred. No. 2.7e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SLFTPGAKONIQLINT 20
Db 80 SLFTPGAKONIQLINT 95

RESULT 4
Q917R8 PRELIMINARY; PRT; 169 AA.
ID Q917R8
AC Q917R8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=hv432-1;
DR EMBL; AF416526; AAL10426.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 169
SQ SEQUENCE 169 AA; 18315 MW; 055C30C48763696A CRC64;

Query Match 80.0%; Score 16; DB 12; Length 169;
Best Local Similarity 100.0%; Pred. No. 2.7e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SLFTPGAKONIQLINT 20
Db 80 SLFTPGAKONIQLINT 95

RESULT 5
Q917R9 PRELIMINARY; PRT; 169 AA.
ID Q917R9
AC Q917R9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
```

```

[1]
RN SEQUENCE FROM N.A.
RC STRAIN=hv431-1;
DR EMBL; AF416525; AAL10425.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 169
SQ SEQUENCE 169 AA; 18341 MW; 2D29BB2C3F5A1B00 CRC64;

Query Match 80.0%; Score 16; DB 12; Length 169;
Best Local Similarity 100.0%; Pred. No. 2.7e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SLFTPGAKONIQLINT 20
Db 80 SLFTPGAKONIQLINT 95

RESULT 6
Q917R8 PRELIMINARY; PRT; 170 AA.
ID Q917R8
AC Q917R8;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UK22395;
RA Candotti D., Richetin A., Cant B., Temple J., Sims C., Reeves I.,
RA Barbara J.J., Allain J.-P.;
RT "Evaluation of a TMA-based HCV/HIV-1 RNA duplex assay for screening
RT individual blood donations: A comparison with the minipool testing
RT systems."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF488361; AAL92152.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 170
SQ SEQUENCE 170 AA; 18515 MW; 6F72702C683402D1 CRC64;

Query Match 80.0%; Score 16; DB 12; Length 170;
Best Local Similarity 100.0%; Pred. No. 2.7e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SLFTPGAKONIQLINT 20
Db 69 SLFTPGAKONIQLINT 84

RESULT 7
Q917B9 PRELIMINARY; PRT; 154 AA.
ID Q917B9
AC Q917B9;
```

DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ACL-34;
 RA Lyra A.C., Fan X., Di Bisceglie A.M.;
 RT "Evolution of hepatitis C viral quasiespecies after liver
 RT transplantation."
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF422279; AAL24592.1; -.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 154
 SQ SEQUENCE 154 AA; 16818 MW; 5D6F576F7273248B CRC64;

Query Match 75.0%; Score 15; DB 12; Length 154;
 Best Local Similarity 100.0%; Pred. No. 2.8e-08;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 LFTPGAKONIQLINT 20
 |||||
 Db 117 LFTPGAKONIQLINT 131

RESULT 8

ID Q917B7 PRELIMINARY; PRT; 154 AA.
 AC Q917B7;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ACL-36;
 RA Lyra A.C., Fan X., Di Bisceglie A.M.;
 RT "Evolution of hepatitis C viral quasiespecies after liver
 RT transplantation."
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF422281; AAL24594.1; -.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 154
 SQ SEQUENCE 154 AA; 16864 MW; BC7CF5E598E8FE6 CRC64;

Query Match 75.0%; Score 15; DB 12; Length 154;
 Best Local Similarity 100.0%; Pred. No. 2.8e-08;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 LFTPGAKONIQLINT 20
 |||||

Db 117 LFTPGAKONIQLINT 131

RESULT 9

ID Q8QQQ8 PRELIMINARY; PRT; 154 AA.
 AC Q8QQQ8;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE E1/E2 protein (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ACL-262;
 RA Lyra A.C., Fan X., Di Bisceglie A.M.;
 RT "Evolution of hepatitis C viral quasiespecies after liver
 RT transplantation."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF497280; AAM19536.1; -.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 154
 SQ SEQUENCE 154 AA; 16818 MW; 8F6DF7E758732488 CRC64;

Query Match 75.0%; Score 15; DB 12; Length 154;
 Best Local Similarity 100.0%; Pred. No. 2.8e-08;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 LFTPGAKONIQLINT 20
 |||||
 Db 117 LFTPGAKONIQLINT 131

RESULT 10

ID Q917B5 PRELIMINARY; PRT; 154 AA.
 AC Q917B5;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ACL-38;
 RA Lyra A.C., Fan X., Di Bisceglie A.M.;
 RT "Evolution of hepatitis C viral quasiespecies after liver
 RT transplantation."
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF422283; AAL24596.1; -.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 154
 SQ SEQUENCE 154 AA; 16709 MW; 231B0179377F187B CRC64;

Query Match 75.0%; Score 15; DB 12; Length 154;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LFTPGAKONIOLINT 20
DB 117 LFTPGAKONIOLINT 131

RESULT 11

Q917D4 PRELIMINARY; PRT; 154 AA.
AC Q917D4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ACL-19;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Evolution of hepatitis C viral quasiespecies after liver transplantation."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF422264; AAL24577.1; -;
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 154
SQ SEQUENCE 154 AA; 16818 MW; 8F6DF7E758732488 CRC64;

Query Match 75.0%; Score 15; DB 12; Length 154;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LFTPGAKONIOLINT 20
DB 117 LFTPGAKONIOLINT 131

RESULT 12

Q917D7 PRELIMINARY; PRT; 154 AA.
AC Q917D7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ACL-16;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Evolution of hepatitis C viral quasiespecies after liver transplantation."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF422261; AAL24574.1; -;
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.

DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 154
SQ SEQUENCE 154 AA; 16818 MW; 8F6DF7E758732488 CRC64;

Query Match 75.0%; Score 15; DB 12; Length 154;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LFTPGAKONIOLINT 20
DB 117 LFTPGAKONIOLINT 131

RESULT 13

Q800Q3 PRELIMINARY; PRT; 154 AA.
AC Q800Q3;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE E1/E2 protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ACL-267;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Evolution of hepatitis C viral quasiespecies after liver transplantation."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF497285; AAMI9541.1; -;
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 154
SQ SEQUENCE 154 AA; 16864 MW; BC7CF5E55858FE6 CRC64;

Query Match 75.0%; Score 15; DB 12; Length 154;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LFTPGAKONIOLINT 20
DB 117 LFTPGAKONIOLINT 131

RESULT 14

Q800Q4 PRELIMINARY; PRT; 154 AA.
AC Q800Q4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE E1/E2 protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ACL-266;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;

```

RT "Evolution of hepatitis C viral quasiespecies after liver
RT transplantedation.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF497284; AAM19540.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 154
SQ SEQUENCE 154 AA; 16799 MW; F10F576F726B1878 CRC64;

Query Match 75.0%; Score 15; DB 12; Length 154;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LFTPGAKQNIQLINT 20
DB 117 LFTPGAKQNIQLINT 131

RESULT 15
Q917D0
ID Q917D0 PRELIMINARY; PRT; 154 AA.
AC Q917D0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ACU-23;
RA Lyza A.C., Fan X., Di Bisceglie A.M.;
RT "Evolution of hepatitis C viral quasiespecies after liver
RT transplantedation.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF422268; AAL24581.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 154
SQ SEQUENCE 154 AA; 16818 MW; 8F6DF7E758732488 CRC64;

Query Match 75.0%; Score 15; DB 12; Length 154;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LFTPGAKQNIQLINT 20
DB 117 LFTPGAKQNIQLINT 131

Search completed: November 21, 2003, 21:08:11
Job time : 24.45 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:22 ; Search time 32.1 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-73

Perfect score: 20

Sequence: 1 QNIQLINTNGQWHINSTALN 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A Geneseq_19Jun03.*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	20	100.0	20	AAO18702	Hepatitis C virus
2	10	50.0	20	AA184434	Hepatitis C virus
3	10	50.0	20	AA190996	HCV E2 peptide E2-
4	10	50.0	30	AAU84624	HCV HepC1a segment
5	10	50.0	30	AAU84625	HCV HepC1a segment
6	10	50.0	57	AA193877	Polypeptide p380.L
7	10	50.0	91	ABB77256	HCV bait polypepti
8	10	50.0	97	AAO22058	Hepatitis C virus
9	10	50.0	111	AAO22061	Hepatitis C virus

10	10	50.0	144	14	AA183998	HC-J1 E2/NS1 prote
11	10	50.0	174	14	AA183441	Sequence of glycop
12	10	50.0	278	14	AA183997	Th E2/NS1 protein.
13	10	50.0	305	18	AAW00929	Recombinant HCV E2
14	10	50.0	319	15	AAW45330	Anti-HCV antibody
15	30	50.0	333	14	AA1840118	HGH-HCV-E2 fusion
16	10	50.0	337	16	AA179217	pHCV351-encoded AP
17	10	50.0	350	23	AA189891	Hepatitis C virus
18	10	50.0	363	24	AA1832880	Hepatitis C virus
19	10	50.0	363	24	ABP55567	Hepatitis C virus
20	10	50.0	367	14	AA1840115	APP-HCV-E2 fusion
21	10	50.0	367	16	AA179218	pHCV167-encoded pr
22	10	50.0	377	16	AA179226	pHCV422-encoded AP
23	10	50.0	397	16	AA179220	pHCV419-encoded AP
24	10	50.0	399	14	AA1840117	HGH-HCV-E2 fusion
25	10	50.0	402	14	AA1834439	Sequence of glycop
26	10	50.0	409	14	AA1833995	H77 E2/NS1 protein
27	10	50.0	410	16	AA179227	pHCV423-encoded AP
28	10	50.0	417	16	AA179228	pHCV424-encoded AP
29	10	50.0	434	16	AA179219	pHCV418-encoded AP
30	10	50.0	441	16	AA179230	pHCV429-encoded AP
31	10	50.0	447	16	AA179239	pHCV425-encoded AP
32	10	50.0	453	16	AA179225	pHCV421-encoded AP
33	10	50.0	467	13	AA1831575	HCV CKS-3'ENV - pH
34	10	50.0	467	14	AA1833642	HCV CKS-3'ENV reco
35	10	50.0	467	14	AA1833584	HCV CKS-3'ENV reco
36	10	50.0	467	14	AA1833604	HCV CKS-3'ENV reco
37	10	50.0	467	22	AA1869003	HCV recombinant an
38	10	50.0	490	16	AA179224	pHCV420-encoded AP
39	10	50.0	502	16	AA1867589	Hepatitis C virus
40	10	50.0	513	13	AA1824086	NANB hepatitis vir
41	10	50.0	527	13	AA1825136	HCV polypeptide 2.
42	10	50.0	537	14	AA1840114	APP-HCV-E2 fusion
43	10	50.0	617	13	AA1821578	HCV CKS-ENV - pHCV
44	10	50.0	617	14	AA1833645	HCV CKS-ENV recomb
45	10	50.0	617	14	AA1833587	HCV CKS-ENV recomb

ALIGNMENTS

RESULT 1

AAO18702

ID AAO18702 standard; Peptide; 20 AA.

XX AC AAO18702;

XX DT 24-OCT-2002 (first entry)

XX Hepatitis C virus E2 protein derived peptide E2-69.

XX Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide;
XX immunostimulant; vaccine.

XX Hepatitis C virus.

XX WO200255548-A2.

XX PD 18-JUL-2002.

XX PF 11-JAN-2002; 2002WO-EP00219.

XX PR 11-JAN-2001; 2001US-260699P.

XX PR 30-AUG-2001; 2001US-315768P.

XX (INNO-) INNOGENETICS NV.

XX Maertens G, Bosman F, Buysse M;

XX WPI; 2002-599657/64.

XX New therapeutic vaccine compositions comprising at least one purified
PT recombinant hepatitis C virus (HCV) single or specific oligomeric

PT recombinant envelope protein E1 or E2, useful for immunizing humans
 PT from HCV infection
 XX
 PS Claim 4; Page 226; 243pp; English.
 XX
 CC The present invention relates to new therapeutic vaccine compositions for
 CC inducing hepatitis C virus (HCV)-specific antibodies, comprising a
 CC composition containing at least one purified recombinant HCV single or
 CC specific oligomeric recombinant envelope proteins selected from an E1 and
 CC an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are
 CC useful for inducing HCV-specific antibodies or for immunising humans
 CC against HCV. The recombinant HCV E1 and/or E2 proteins are useful as
 CC vaccines or therapeutics, in HCV screening and confirmatory antibody
 CC tests, for raising antibodies, in the preparation of medicament, and for
 CC in vitro monitoring of HCV disease or prognosing the response to
 CC treatment of patients suffering from HCV infection. The present sequence
 CC is a peptide derived from the proteins of the invention.
 XX
 SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 23; Length 20;
 Best Local Similarity 100.0%; Pred. No. 2.4e-14;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNIQLINTNGQWHINSTALN 20
 DB 1 QNIQLINTNGQWHINSTALN 20

RESULT 2
 AAR84434
 ID AAR84434 standard; peptide; 20 AA.
 XX
 AC AAR84434;

DT 06-JAN-1997 (first entry)
 XX
 DE Hepatitis C virus peptide NS1-5* (residues 409-428).
 XX
 KW Hepatitis C virus; HCV; immunogen; non-structural region; E2 region;
 KW immunodominant; T cell epitope; vaccine.
 XX
 OS Hepatitis C virus.

XX
 PN WO9512677-A2.
 XX
 PD 11-MAY-1995.
 XX
 PF 28-OCT-1994; 94WO-EP03555.
 XX
 PR 04-NOV-1993; 93EP-0402718.
 XX
 PA (INNO-) INNOGENETICS NV.
 XX
 PI Deleys R, Leroux-Roels G, Maertens G;
 XX
 DR WPI; 1995-193822/25.

XX
 PT Hepatitis C Virus immunogenic polypeptide contg. a T-cell
 PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in
 PT production of vaccines, therapeutic agents, etc.
 XX
 PS Claim 16; Page 66; 105pp; English.
 XX
 CC Polypeptides comprising 8-32 amino acids from the HCV E2 region
 CC sequence spanning positions 397-428 and containing a T-cell
 CC stimulating epitope are used in HCV immunogenic compositions.
 CC The present sequence is a specific example of a T-cell epitope-
 CC containing peptide which is preferred for use in the compositions.

XX
 SQ Sequence 20 AA;
 Query Match 50.0%; Score 10; DB 16; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.00073;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNIQLINTNG 10
 DB 1 QNIQLINTNG 10

RESULT 3
 AAR90996
 ID AAR90996 standard; peptide; 20 AA.
 XX
 AC AAR90996;

DT 25-SEP-1996 (first entry)
 XX
 DE HCV E2 peptide E2-69 for competition studies.

XX
 KW HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;
 KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.
 XX
 OS Synthetic.

XX
 PN WO9604385-A2.
 XX
 PD 15-FEB-1996.

XX
 PF 31-JUL-1995; 95WO-EP03031.
 XX
 PR 29-JUL-1994; 94EP-0870132.

XX
 PA (INNO-) INNOGENETICS NV.

XX
 PI Bosman F, Buyse M, De Martynoff G, Maertens G;
 XX
 DR WPI; 1996-129401/13.

XX
 PT Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope
 PT proteins - in presence of disulphide bond cleavage agent, to
 PT produce proteins suitable for direct use in vaccines or diagnostic
 PT assays of HCV

XX
 PS Claim 29; Page 67; 146pp; English.

XX
 CC AAR90965-R90980 and AAR90995-R91015 represent synthetic hepatitis C
 CC virus (HCV) E1 and E2 peptides used in competition studies. This
 CC sequence represents a synthetic E2 peptide, and corresponds to residues
 CC 409-428 of the E2 protein sequence. These sequences are useful for in
 CC vitro monitoring of HCV disease, or prognosis of the response to
 CC interferon treatment of patients suffering from HCV infection. These
 CC sequences compete with the proteins produced by AAT12704-T12709 and
 CC AAT12961-T12974, which are included in vectors for the production of
 CC recombinant E1, E2, and E1/E2. The recombinant proteins can then be
 CC isolated and purified by carrying out a disulphide bond cleavage, or a
 CC reduction step with a disulphide bond cleavage agent, after lysis of
 CC recombinant host cells. The constructs containing the purified HCV
 CC envelope proteins can be used for vaccinating humans against HCV, for in
 CC vitro detection of HCV antibodies in a sample, and in a serotyping assay
 CC for detecting one or more serological types of HCV present in a
 CC biological sample. The constructs can also be immobilised on a solid
 CC substrate and incorporated into a reversed phase hybridisation assay for
 CC determining the presence or the genotype of HCV. The new purification
 CC method preserves the conformation of the recombinantly expressed E1, E2
 CC and E1/E2, and eliminates contaminating proteins. Antigens isolated
 CC using this method are more reactive with human sera than those isolated
 CC by known techniques.

XX
 SQ Sequence 20 AA;

Query Match 50.0%; Score 10; DB 17; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.00073;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNIQLINTNG 10
 |||||
 Db 1 QNIQLINTNG 10

RESULT 4
 AAU84624
 ID AAU84624 standard; Peptide; 30 AA.
 AC AAU84624;

XX 08-MAY-2002 (first entry)

DE HCV HepC1a segment 27.

XX Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
 KW viral infection; human immunodeficiency virus; melanoma;
 KW bacterial infection; Salmonella; Legionella; parasitic infection;
 KW Trypanosoma; Toxoplasma; Giardia.

XX Hepatitis C virus.

OS WO200190197-A1.

XX 29-NOV-2001.

XX 25-MAY-2001; 2001WO-AU00622.

XX 26-MAY-2000; 2000AU-0007761.

XX (AUSU) UNIV AUSTRALIAN NAT.

XX Thomson SA, Ramshaw IA;

XX WPI; 2002-147575/19.

DR N-PSDB; ABK36462.

XX New synthetic polypeptides having several different segments of at
 PT least one parent polypeptide linked together differently compared to
 PT the linkage in the parent polypeptide, for inducing immune response
 PT against a pathogen or cancer -

XX Example 2; Fig 26; 364pp; English.

XX The invention relates to a new synthetic polypeptide (I) comprising
 CC several different segments of at least one parent polypeptide linked
 CC together in a different relationship relative to their linkage in the
 CC parent polypeptide to impede, abrogate or otherwise alter at least one
 CC function associated with the parent polypeptide and for inducing an
 CC immune response against a pathogen or cancer. Also included are a
 CC synthetic polynucleotide encoding and a computer system for
 CC designing the synthetic polypeptides. The synthetic polypeptides and
 CC polynucleotides are referred to as a savine. The synthetic polypeptide is
 CC useful for modulating immune responses preferably directed against a
 CC pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix,
 CC colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone
 CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.
 CC Compositions comprising the polypeptide may be used in the treatment or
 CC prophylaxis against viral (such as infections caused by HIV (human
 CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
 CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
 CC (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,
 CC Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic
 CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
 CC Trypanosoma, Toxoplasma and Giardia) infections. The present
 CC sequence is a peptide derived from a parent protein used to
 CC construct a savine of the invention.

XX Sequence 30 AA;

Query Match 50.0%; Score 10; DB 23; Length 30;
 Best Local Similarity 100.0%; Pred. No. 0.0011;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNIQLINTNG 10
 |||||
 Db 21 QNIQLINTNG 30

RESULT 5
 AAU84625
 ID AAU84625 standard; Peptide; 30 AA.
 XX AAU84625;

XX 08-MAY-2002 (first entry)

DE HCV HepC1a segment 28.

XX Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
 KW viral infection; human immunodeficiency virus; melanoma;
 KW bacterial infection; Salmonella; Legionella; parasitic infection;
 KW Trypanosoma; Toxoplasma; Giardia.

XX Hepatitis C virus.

XX WO200190197-A1.

XX 29-NOV-2001.

XX 25-MAY-2001; 2001WO-AU00622.

XX 26-MAY-2000; 2000AU-0007761.

XX (AUSU) UNIV AUSTRALIAN NAT.

XX Thomson SA, Ramshaw IA;

XX WPI; 2002-147575/19.

DR N-PSDB; ABK36463.

XX New synthetic polypeptides having several different segments of at
 PT least one parent polypeptide linked together differently compared to
 PT the linkage in the parent polypeptide, for inducing immune response
 PT against a pathogen or cancer -

XX Example 2; Fig 26; 364pp; English.

XX The invention relates to a new synthetic polypeptide (I) comprising
 CC several different segments of at least one parent polypeptide linked
 CC together in a different relationship relative to their linkage in the
 CC parent polypeptide to impede, abrogate or otherwise alter at least one
 CC function associated with the parent polypeptide and for inducing an
 CC immune response against a pathogen or cancer. Also included are a
 CC synthetic polynucleotide encoding and a computer system for
 CC designing the synthetic polypeptides. The synthetic polypeptides and
 CC polynucleotides are referred to as a savine. The synthetic polypeptide is
 CC useful for modulating immune responses preferably directed against a
 CC pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix,
 CC colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone
 CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.
 CC Compositions comprising the polypeptide may be used in the treatment or
 CC prophylaxis against viral (such as infections caused by HIV (human
 CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
 CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
 CC (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,
 CC Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic
 CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
 CC Trypanosoma, Toxoplasma and Giardia) infections. The present
 CC sequence is a peptide derived from a parent protein used to
 CC construct a savine of the invention.

XX Sequence 30 AA;

Query Match 50.0%; Score 10; DB 23; Length 30;
 Best Local Similarity 100.0%; Pred. No. 0.0011;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNIQLINTNG 10
|||||
Db 6 QNIQLINTNG 15

RESULT 6
AAR33877
ID AAR33877 standard; peptide; 57 AA.

XX AC AAR33877;
XX DT 25-MAR-2003 (updated)
XX DT 19-JUL-1993 (first entry)

XX DE Polypeptide p380.LG comprising HCV viral antigen.

XX KW Hepatitis C virus; NANBH; assay; antibody; p380-JH1; p380-J; p380LG;
XX KW p408.

XX OS Synthetic.

XX PN WO9306247-A1.

XX PD 01-APR-1993.

XX PF 16-SEP-1992; 92WO-US07813.

XX PR 16-SEP-1991; 91US-0760292.

XX PA (ABBO) ABBOTT LAB.

XX PI Lesniewski RR, Leung TK;

XX DR WPI; 1993-117563/14.

XX PT Assay for detecting presence of antibody to hepatitis C viral
XX PT antigen - by contacting sample with polypeptide contg. at least
XX PT one epitope of virus antigen

XX PS Disclosure; Page 14; 63pp; English.

XX CC The synthetic peptide p380.LG represents amino acid residues 380-436 of
XX CC the hepatitis C viral antigen. The peptide may be used in an assay to
XX CC detect antibodies to HCV and thus to diagnose chronic HCV infection.

XX CC See also AAR33861-87.

XX CC (Updated on 25-MAR-2003 to correct PN field.)

XX SQ Sequence 57 AA;

Query Match 50.0%; Score 10; DB 14; Length 57;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNIQLINTNG 10
|||||
Db 30 QNIQLINTNG 39

RESULT 7
ABB77256
ID ABB77256 standard; Protein; 91 AA.

XX AC ABB77256;

XX DT 28-JUN-2002 (first entry)

XX DE HCV bait polypeptide 4.

XX KW SID; selected interacting domain; HCV; hepatitis C virus;
XX KW liver disease; liver cancer; virucide; hepatotropic; antiinflammatory;
XX KW antibacterial.

XX OS Hepatitis C virus strain H77.

XX PN EP1178116-A1.

XX PD 06-FEB-2002.

XX PF 03-AUG-2000; 2000EP-0402225.

XX PR 03-AUG-2000; 2000EP-0402225.

XX PA (HYBR-) HYBRIGENICS SA.

XX PI Legrain P, Whiteside S, Wojcik J;

XX DR WPI; 2002-208115/27.

XX DR N-PSDB; ABL55588.

XX PT New selected interacting domain polypeptides and polynucleotides,
XX PT useful for treating or preventing infections or pathologies caused by
XX PT hepatitis C virus (HCV) or those linked to HCV infection

XX PS Claim 26; SEQ ID 80; 61pp + sequence listing; English.

XX CC The invention relates to nucleic acids encoding polypeptides which are
XX CC termed SID polypeptides (selected interacting domain). These polypeptides
XX CC are the final products of a double selection method involving a first
XX CC step of selection of Hepatitis C virus (HCV)-derived polynucleotides
XX CC through a two-hybrid system, and a second selection step involving an
XX CC alignment between the different polynucleotides selected at the first
XX CC step. The activity of polypeptides of the invention may be described as,
XX CC virucide, hepatotropic, antiinflammatory and antibacterial. The
XX CC polypeptide, polynucleotide and compositions comprising them are useful
XX CC for treating or preventing viral or a bacterial infection, specifically
XX CC infections or pathologies caused by HCV, or those pathologies linked to
XX CC HCV infection. These may include liver disease and liver cancer. The
XX CC current sequence represents a HCV bait polypeptide.

XX CC Note: The sequence data for this patent is not represented in the
XX CC specification, but is based on sequence information supplied by the
XX CC European Patent Office.

XX SQ Sequence 91 AA;

Query Match 50.0%; Score 10; DB 23; Length 91;
Best Local Similarity 100.0%; Pred. No. 0.0029;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNIQLINTNG 10
|||||
Db 30 QNIQLINTNG 39

RESULT 8
AAO22058
ID AAO22058 standard; Protein; 97 AA.

XX AC AAO22058;

XX DT 25-SEP-2002 (first entry)

XX DE Hepatitis C virus related protein sequence #11.

XX KW Hypervariable area; Korean hepatitis C virus.

XX OS Unidentified.

XX PN KR301795-B.

XX PD 22-OCT-2001.

XX PF 02-JUN-1993; 93KR-0009915.

XX PR 02-JUN-1993; 93KR-0009915.

XX PA (GLDS) LG CHEM LTD.
 XX PI Cho JM, Park YW;
 XX PT WPI; 2002-398847/43.
 XX DR N-PSDB; AAL40691.
 XX PT Nucleotide sequence and amino acid sequence of hypervariable area of
 XX PT Korean hepatitis C virus -
 XX PS Disclosure; Fig 10; 35pp; Korean.
 XX CC The invention relates to a nucleotide sequence and amino acid sequence of
 CC a hypervariable area of the Korean hepatitis C virus. This sequence
 CC represents a protein encoded by DNA relating to the hypervariable area of
 CC the hepatitis C virus of the invention.
 XX SQ Sequence 97 AA;
 Query Match 50.0%; Score 10; DB 23; Length 97;
 Best Local Similarity 100.0%; Pred. No. 0.0031;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 QNIQLINTNG 10
 Db 85 QNIQLINTNG 94
 AAO22061
 ID AAO22061 standard; Protein; 111 AA.
 XX AC AAO22061;
 XX DT 25-SEP-2002 (first entry)
 XX DE Hepatitis C virus related protein sequence #14.
 XX KW Hypervariable area; Korean hepatitis C virus.
 XX OS Unidentified.
 XX FN KR301795-B.
 XX PD 22-OCT-2001.
 XX PF 02-JUN-1993; 93KR-0009915.
 XX PR 02-JUN-1993; 93KR-0009915.
 XX PA (GLDS) LG CHEM LTD.
 XX PI Cho JM, Park YW;
 XX DR WPI; 2002-398847/43.
 XX DR N-PSDB; AAL40691.
 XX PT Nucleotide sequence and amino acid sequence of hypervariable area of
 PT Korean hepatitis C virus -
 XX PS Disclosure; Fig 13; 35pp; Korean.
 XX CC The invention relates to a nucleotide sequence and amino acid sequence of
 CC a hypervariable area of the Korean hepatitis C virus. This sequence
 CC represents a protein encoded by DNA relating to the hypervariable area of
 CC the hepatitis C virus of the invention.
 XX SQ Sequence 111 AA;
 Query Match 50.0%; Score 10; DB 23; Length 111;
 Best Local Similarity 100.0%; Pred. No. 0.0035;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
 Db 85 QNIQLINTNG 94
 RESULT 10
 AAR33998
 ID AAR33998 standard; Protein; 144 AA.
 XX AC AAR33998;
 XX DT 25-MAR-2003 (updated)
 XX DT 26-JUL-1993 (first entry)
 XX DE HC-J1 E2/NS1 protein.
 XX KW Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
 KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
 KW domain; immunological; cross-reactive; envelope protein; vaccine;
 KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.
 XX OS Synthetic.
 XX FN WO9306126-A1.
 XX PD 01-APR-1993.
 XX PF 11-SEP-1992; 92WO-US07683.
 XX PR 13-SEP-1991; 91US-0759575.
 XX PA (CHIR) CHIRON CORP.
 XX PI Houghton M, Weiner AJ;
 XX DR WPI; 1993-117468/14.
 XX PT Immuno-reactive hepatitis C virus polypeptide compsns. - contg.
 PT at least 2 sequences from the first variable domain of distinct
 PT HCV isolates
 XX PS Disclosure; Fig 3; 106pp; English.
 XX CC The sequences given in AAR33992-002 represent a portion of the E2/NS1
 CC protein encoded by group I and group II HCV isolates, from amino acid
 CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
 CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
 CC 30 amino acids which shows large variation between nearly all isolates.
 CC This is an important immunoreactive domain. This putative envelope
 CC glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera
 CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
 CC flaviviruses, both of which confer protective immunity in hosts
 CC vaccinated with these polypeptides. It has been discovered that a
 CC number of important HCV epitopes vary among viral isolates and that
 CC these epitopes can be mapped to specific domains. This meant that
 CC immunologically cross-reactive polypeptides which focus on variable
 CC rather than constant domains can be produced. See also AAO39134-48
 CC and AAR33982-91.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX SQ Sequence 144 AA;
 Query Match 50.0%; Score 10; DB 14; Length 144;
 Best Local Similarity 100.0%; Pred. No. 0.0044;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 QNIQLINTNG 10
 Db 40 QNIQLINTNG 49
 RESULT 11

```

AAR34441
ID AAR34441 standard; Protein; 174 AA.
XX AC AAR34441;
XX DT 25-MAR-2003 (updated)
XX DT 09-AUG-1993 (first entry)
XX DE Sequence of glycoprotein E2/NS1 in clone J1(JM).
XX KW Hepatitis C virus; envelope protein; glycoprotein; E2/NS1;
XX KW diagnostic reagent.
XX OS Hepatitis C virus.
XX PN EP537626-A1.
XX PD 21-APR-1993.
XX PF 08-OCT-1992; 92EP-0117191.
XX PR 08-OCT-1991; 91JP-0260824.
XX PA (NAHE-) NAT INST OF HEALTH.
XX PI Harada S, Honda Y, Miyamura T, Saito I;
XX DR WPI; 1993-127516/16.
XX DR N-PSDB; AAQ40333.
XX PT Diagnostic reagent for hepatitis C virus - comprises second
XX PT envelope protein or first non-structural protein encoded by HCV
XX PT gene and has sugar chain
XX PS Claim 2; Pages 40-41; 58pp; English.
XX CC Glycoprotein E2/NS1 is derived from the second envelope protein or
XX CC first non-structural protein encoded by the genome of HCV. The
XX CC nucleic acid is extracted from the serum of the patient of hepatitis
XX CC C. The serum is pref. mixed with transfer RNA (tRNA) as a carrier
XX CC of virus RNA. As a technique of cloning cDNA from the nucleic acid,
XX CC it is preferred to use polymerase chain reaction method. In the
XX CC reaction, any commercially available random primers or synthesized
XX CC DNA having a base sequence similar to that of primer AS1 may be used
XX CC as a primer. Representative examples of sense primers include S1.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 174 AA;
Query Match 50.0%; Score 10; DB 14; Length 174;
Best Local Similarity 100.0%; Pred. No. 0.0052;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QNIQLINTNG 10
Db 70 QNIQLINTNG 79
RESULT 12
AAR33997
ID AAR33997 standard; Protein; 278 AA.
XX AC AAR33997;
XX DT 25-MAR-2003 (updated)
XX DT 26-JUL-1993 (first entry)
XX DE Th E2/NS1 protein.
XX KW Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
XX KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
XX KW domain; immunological; cross-reactive; envelope protein; vaccine;
XX KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.
XX
XX Synthetic.
XX OS W09306126-A1.
XX PN 01-APR-1993.
XX PD 11-SEP-1992; 92WO-US07683.
XX PF 13-SEP-1991; 91US-0759575.
XX PR (CHIR ) CHIRON CORP.
XX PA Houghton M, Weiner AJ;
XX PI WPI; 1993-117468/14.
XX DR Immuno-reactive hepatitis C virus polypeptide compsns. - contg.
XX DR at least 2 sequences from the first variable domain of distinct
XX PT HCV isolates
XX PT Disclosure; Fig 3; 106pp; English.
XX
XX The sequences given in AAR33992-002 represent a portion of the E2/NS1
XX CC protein encoded by group I and group II HCV isolates, from amino acid
XX CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
XX CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
XX CC 30 amino acids which shows large variation between nearly all isolates.
XX CC This is an important immunoreactive domain. This putative envelope
XX CC glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera
XX CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
XX CC flaviviruses, both of which confer protective immunity in hosts
XX CC vaccinated with these polypeptides. It has been discovered that a
XX CC number of important HCV epitopes vary among viral isolates and that
XX CC these epitopes can be mapped to specific domains. This meant that
XX CC immunologically cross-reactive polypeptides which focus on variable
XX CC rather than constant domains can be produced. See also AAQ39134-48
XX CC and AAR33982-91.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 278 AA;
Query Match 50.0%; Score 10; DB 14; Length 278;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QNIQLINTNG 10
Db 40 QNIQLINTNG 49
RESULT 13
AAW00929
ID AAW00929 standard; Protein; 305 AA.
XX AC AAW00929;
XX DT 04-NOV-1997 (first entry)
XX DE Recombinant HCV E2 antigen.
XX KW HCV; E2; antigen; non-secretor gene; protein secretion; vaccine;
XX KW plasmid 577.
XX OS Chimeric hepatitis C virus;
XX OS Chimeric synthetic.
XX PH Key Location/Qualifiers
XX FT Peptide 1..19
XX FT Cleavage-site 20 /label= Sig_peptide
XX FT /note= "mammalian secretion signal
XX FT peptidase cleavage site"

```



```

XX  New plasmid pHCV-162 is a mammalian expression systems for HCV
PT  proteins - useful for diagnosing HCV infection and as vaccines
PT  for preventing HCV infection
XX
XX  Claim 10; Page 81-82; 100pp; English.
XX
XX  A sequence coding for the HCV E2 protein (amino acids 384-684 from
CC  HCV-infected chimpanzee CO isolate, see AAR40119) was generated using
CC  PCR. An EcoRI site was used to attach a synthetic oligonucleotide
CC  encoding the Human Growth Hormone secretion signal at the 5'-end of
CC  the HCV sequence. The resultant plasmid was designated pHCV-170.
CC  AAR40118 is the sequence of the HGH-HCV-E2 fusion protein expressed
CC  by vector pHCV-170. See AAR47192-Q47196.
CC  (Updated on 25-MAR-2003 to correct PN field.)
XX
XX  SQ Sequence 333 AA;
    Query Match          50.0%; Score 10; DB 14; Length 333;
    Best Local Similarity 100.0%; Pred. No. 0.0095;
    Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    QY 1 QNIQLINTNG 10
    DB 58 QNIQLINTNG 67

```

Search completed: November 21, 2003, 20:58:02
Job time : 32.15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.6 Seconds
(without alignments)
79.832 Million cell updates/sec

Title: US-09-973-025-73

Perfect score: 20

Sequence: 1 QNIQLINTNGQWHINSTALN 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents, AA.*

- 1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-08-612-973-73
2	20	100.0	20	3	US-08-927-597-73
3	10	50.0	20	4	US-08-635-886C-47
4	10	50.0	52	1	US-08-262-037-94
5	10	50.0	57	5	PCT-US92-07813-17
6	10	50.0	139	3	US-08-444-818-171
7	10	50.0	139	3	US-08-444-818-174
8	10	50.0	144	1	US-08-440-103-20
9	10	50.0	144	1	US-08-440-542-20
10	10	50.0	144	1	US-08-231-368-20
11	10	50.0	144	1	US-08-440-210-20
12	10	50.0	144	4	US-09-046-604-20
13	10	50.0	174	1	US-08-460-806-19
14	10	50.0	174	1	US-08-325-630-19
15	10	50.0	180	2	US-08-483-695-40
16	10	50.0	180	2	US-07-965-285-40
17	10	50.0	180	2	US-08-487-231-40
18	10	50.0	180	3	US-09-201-912-40
19	10	50.0	278	1	US-08-440-103-15
20	10	50.0	278	1	US-08-440-542-15
21	10	50.0	278	1	US-08-231-368-15
22	10	50.0	278	1	US-08-440-210-15
23	10	50.0	278	4	US-09-046-604-15
24	10	50.0	305	3	US-08-478-073-2
25	10	50.0	333	1	US-08-453-552-12
26	10	50.0	333	2	US-08-710-637-12
27	10	50.0	333	5	PCT-US93-00907-12

28	10	50.0	337	1	US-08-188-281B-7	Sequence 7, Appli
29	10	50.0	337	5	PCT-US94-07280-7	Sequence 7, Appli
30	10	50.0	337	5	PCT-US95-01087-7	Sequence 7, Appli
31	10	50.0	339	3	US-08-444-818-150	Sequence 150, App
32	10	50.0	339	3	US-08-444-818-158	Sequence 158, App
33	10	50.0	367	1	US-08-188-281B-9	Sequence 9, Appli
34	10	50.0	367	2	US-08-453-552-6	Sequence 6, Appli
35	10	50.0	367	2	US-08-710-637-6	Sequence 6, Appli
36	10	50.0	367	5	PCT-US93-00907-6	Sequence 6, Appli
37	10	50.0	367	5	PCT-US94-07280-9	Sequence 9, Appli
38	10	50.0	367	5	PCT-US95-01087-9	Sequence 9, Appli
39	10	50.0	377	1	US-08-188-281B-17	Sequence 17, Appli
40	10	50.0	377	5	PCT-US94-07280-17	Sequence 17, Appli
41	10	50.0	377	5	PCT-US95-01087-17	Sequence 17, Appli
42	10	50.0	397	1	US-08-188-281B-11	Sequence 11, Appli
43	10	50.0	397	5	PCT-US94-07280-11	Sequence 11, Appli
44	10	50.0	397	5	PCT-US95-01087-11	Sequence 11, Appli
45	10	50.0	399	1	US-08-453-552-10	Sequence 10, Appli

ALIGNMENTS

RESULT 1
US-08-612-973-73
; Sequence 73, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P. C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612.973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-612-973-73
Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QNIQLINTNGQWHINSTALN 20
|||||||

Db 1 QNIQLINTNGQWHINSTALN 20

RESULT 2

US-08-927-597-73
; Sequence 73, Application US/08927597
; Patent No. 6245503
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTINOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-927-597-73

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNGQWHINSTALN 20

Db 1 QNIQLINTNGQWHINSTALN 20

RESULT 3

US-08-635-886C-47
; Sequence 47, Application US/08635886C
; Patent No. 655114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555

; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-47

Query Match 50.0%; Score 10; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.00025;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10

Db 1 QNIQLINTNG 10

RESULT 4

US-08-262-037-94
; Sequence 94, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-262-037-94

Query Match 50.0%; Score 10; DB 1; Length 52;
Best Local Similarity 100.0%; Pred. No. 0.0006;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
Db 2 QNIQLINTNG 11

RESULT 5

PCT-US92-07813-17
; Sequence 17, Application PC/TUS9207813
; GENERAL INFORMATION:
; APPLICANT: LESNIEWSKI, RICHARD R.
; APPLICANT: LEUNG, TAT K.
; TITLE OF INVENTION: HEPATITIS C ASSAY
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES CHAD377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: U.S.A.
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/07813
; FILING DATE: 19920916
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: POREBSKIP, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4767.P3.03
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 57 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US92-07813-17

Query Match 50.0%; Score 10; DB 5; Length 57;
Best Local Similarity 100.0%; Pred. No. 0.00066;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
Db 30 QNIQLINTNG 39

RESULT 6

US-08-444-818-171
; Sequence 171, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA

ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 171:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 139 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: JH23
US-08-444-818-171

Query Match 50.0%; Score 10; DB 3; Length 139;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
Db 110 QNIQLINTNG 119

RESULT 7

US-08-444-818-174
; Sequence 174, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (508)359-3876
TELEFAX: (508)359-3885
INFORMATION FOR SEQ ID NO: 174:
SEQUENCE CHARACTERISTICS:
LENGTH: 139 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE: EC10 (Italy)
INDIVIDUAL ISOLATE: 174

Query Match 50.0%; Score 10; DB 3; Length 139;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
|||||
Db 110 QNIQLINTNG 119

RESULT 8
US-08-440-103-20
; Sequence 20, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,103
FILING DATE: 12-MAY-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
FILING DATE: 13-SEP-1991

ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-103-20

Query Match 50.0%; Score 10; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10

Db 40 QNIQLINTNG 49
|||||

RESULT 9
US-08-440-542-20
; Sequence 20, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,542
FILING DATE: 12-MAY-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
FILING DATE: 13-SEP-1991

ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-542-20

Query Match 50.0%; Score 10; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
|||||
Db 40 QNIQLINTNG 49

RESULT 10
US-08-231-368-20
; Sequence 20, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA

```
;
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-231-368-20

Query Match 50.0%; Score 10; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
Db 40 QNIQLINTNG 49

RESULT 11
US-08-440-210-20
; Sequence 20, Application US/08440210
; Patent No. 5766845
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,210
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
```

```
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-210-20

Query Match 50.0%; Score 10; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
Db 40 QNIQLINTNG 49

RESULT 12
US-09-046-604-20
; Sequence 20, Application US/09046604
; Patent No. 6303292
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/046,604
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-046-604-20

Query Match 50.0%; Score 10; DB 4; Length 144;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
Db 40 QNIQLINTNG 10
```


;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/07/965,285
;; FILING DATE: 18-MAR-1993
;; APPLICATION NUMBER: FR 91 06 882
;; FILING DATE: 06-JUN-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meyers, Kenneth J.
;; REGISTRATION NUMBER: 25,146
;; REFERENCE/DOCKET NUMBER: 05286-0001-00000
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-408-4000
;; TELEFAX: 202-408-4400
;; INFORMATION FOR SEQ ID NO: 40:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 180 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-483-695-40

Query Match 50.0%; Score 10; DB 2; Length 180;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
| | | | | | | |
Db 76 QNIQLINTNG 85

Search completed: November 21, 2003, 21:15:14
Job time : 11.6 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time 19.5 Seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-73

Perfect score: 20

Sequence: 1 QNIQLINTNGWHINSTALN 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	10	US-09-973-025-73
2	20	100.0	20	11	US-09-899-303-73
3	20	100.0	20	11	US-09-995-808-73
4	20	100.0	20	11	US-09-995-860-73
5	20	100.0	20	12	US-09-995-791-73
6	10	50.0	91	10	US-09-921-397-80
7	10	50.0	160	12	US-10-292-129-15
8	10	50.0	350	10	US-09-929-955-4
9	10	50.0	350	14	US-10-104-966-4
10	10	50.0	363	12	US-10-128-587A-97
11	10	50.0	363	15	US-10-128-590-97
12	10	50.0	2894	10	US-09-941-611-23
13	10	50.0	2894	15	US-10-044-995-23
14	10	50.0	3011	9	US-09-742-659-4
15	10	50.0	3011	10	US-09-238-076-20

16	10	50.0	3011	10	US-09-952-572-9	Sequence 9, Appli
17	10	50.0	3011	10	US-09-929-955-1	Sequence 1, Appli
18	10	50.0	3011	10	US-09-747-419-20	Sequence 20, Appl
19	10	50.0	3011	11	US-09-891-894-3	Sequence 3, Appli
20	10	50.0	3011	11	US-09-995-937-20	Sequence 20, Appl
21	10	50.0	3011	11	US-09-917-563-20	Sequence 20, Appl
22	10	50.0	3011	12	US-10-184-150-3	Sequence 1, Appli
23	10	50.0	3011	14	US-10-104-966-1	Sequence 3, Appli
24	10	50.0	3011	15	US-10-259-275-20	Sequence 20, Appl
25	10	50.0	3012	10	US-09-238-076-2	Sequence 2, Appli
26	10	50.0	3012	11	US-09-995-937-2	Sequence 2, Appli
27	10	50.0	3012	11	US-09-917-563-2	Sequence 2, Appli
28	9	45.0	250	10	US-09-952-572-8	Sequence 8, Appli
29	8	40.0	20	10	US-09-973-025-72	Sequence 72, Appl
30	8	40.0	20	11	US-09-899-303-72	Sequence 72, Appl
31	8	40.0	20	11	US-09-995-808-72	Sequence 72, Appl
32	8	40.0	20	11	US-09-995-860-72	Sequence 72, Appl
33	8	40.0	20	12	US-09-995-791-72	Sequence 72, Appl
34	8	40.0	23	16	US-10-318-200-26	Sequence 26, Appl
35	8	40.0	2985	15	US-10-259-275-40	Sequence 40, Appl
36	7	35.0	176	11	US-09-899-046-158	Sequence 158, App
37	7	35.0	176	11	US-09-878-281-158	Sequence 158, App
38	7	35.0	254	10	US-09-407-430-3	Sequence 3, Appli
39	7	35.0	363	10	US-09-407-430-2	Sequence 2, Appli
40	7	35.0	637	12	US-10-187-257-4	Sequence 4, Appli
41	7	35.0	637	12	US-10-265-083-2	Sequence 2, Appli
42	7	35.0	3011	9	US-09-916-359-2	Sequence 2, Appli
43	7	35.0	3011	16	US-10-232-643-6	Sequence 6, Appli
44	6	30.0	23	16	US-10-318-200-28	Sequence 28, Appl
45	6	30.0	24	16	US-10-318-200-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1

US-09-973-025-73
; Sequence 73, Application US/09973025
; Publication No. US20020182706A1

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

BOSMAN, FONS

DE MARTYNOFF, GUY

BUYSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHUYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/973,025

FILING DATE: 10-Oct-2001

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-09-973-025-73

Query Match 100.0%; Score 20; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNGQWHINSTALN 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 QNIQLINTNGQWHINSTALN 20

RESULT 2

US-09-899-303-73
; Sequence 73, Application US/09899303
; Publication No. US20030036110A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/899,303

FILING DATE: 06-Jul-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 73:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 73:

US-09-899-303-73

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNGQWHINSTALN 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 QNIQLINTNGQWHINSTALN 20

Db 1 QNIQLINTNGQWHINSTALN 20

RESULT 3

US-09-995-808-73
; Sequence 73, Application US/09995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:

; APPLICANT: Innogenetics N.V.

; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.

; FILE REFERENCE: 2551-70

; CURRENT APPLICATION NUMBER: US/09/995,808

; CURRENT FILING DATE: 2001-11-29

; NUMBER OF SEQ ID NOS: 122

; SOFTWARE: PatentIn 3.1

; SEQ ID NO 73

; LENGTH: 20

; TYPE: PRT

; ORGANISM: Hepatitis C virus

US-09-995-808-73

Query Match 100.0%; Score 20; DB 11; Length 20;

Best Local Similarity 100.0%; Pred. No. 3.3e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNGQWHINSTALN 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 QNIQLINTNGQWHINSTALN 20

RESULT 4

US-09-995-860-73

; Sequence 73, Application US/09995860

; Publication No. US20030118603A1

; GENERAL INFORMATION:

; APPLICANT: Innogenetics N.V.

; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.

; FILE REFERENCE: 2551-69

; CURRENT APPLICATION NUMBER: US/09/995,860

; CURRENT FILING DATE: 2001-11-29

; NUMBER OF SEQ ID NOS: 122

; SOFTWARE: PatentIn 3.1

; SEQ ID NO 73

; LENGTH: 20

; TYPE: PRT

; ORGANISM: Hepatitis C virus

US-09-995-860-73

Query Match 100.0%; Score 20; DB 11; Length 20;

Best Local Similarity 100.0%; Pred. No. 3.3e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNGQWHINSTALN 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 QNIQLINTNGQWHINSTALN 20

RESULT 5

US-09-995-791-73

; Sequence 73, Application US/09995791

; Publication No. US20030147918A1

; GENERAL INFORMATION:

; APPLICANT: Innogenetics N.V.

; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.

; FILE REFERENCE: 2551-68

; CURRENT APPLICATION NUMBER: US/09/995,791

; CURRENT FILING DATE: 2001-11-29

; NUMBER OF SEQ ID NOS: 122

; SOFTWARE: PatentIn 3.1

; SEQ ID NO 73

```
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-73

Query Match      100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNGQWHINSTALN 20
    |||||
Db 1 QNIQLINTNGQWHINSTALN 20
    |||||

RESULT 6
US-09-921-397-80
; Sequence 80, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A-JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 80
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-80

Query Match      50.0%; Score 10; DB 10; Length 91;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
    |||||
Db 30 QNIQLINTNG 39

RESULT 7
US-10-292-129-15
; Sequence 15, Application US/10292129
; Publication No. US20030148267A1
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Emmett Vance
; APPLICANT: Chung, Raymond Taeyong
; TITLE OF INVENTION: SCREENING ASSAY FOR HEPATITIS C VIRUS
; TITLE OF INVENTION: ANTIVIRAL AGENTS
; FILE REFERENCE: 00786-539001
; CURRENT APPLICATION NUMBER: US/10/292,129
; CURRENT FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: US 60/345,405
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-292-129-15

Query Match      50.0%; Score 10; DB 12; Length 160;
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
    |||||
Db 30 QNIQLINTNG 39

Query Match      100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNGQWHINSTALN 20
    |||||
Db 1 QNIQLINTNGQWHINSTALN 20
    |||||

RESULT 8
US-09-929-955-4
; Sequence 4, Application US/09929955
; Patent No. US20020136740A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.23AUS2
; CURRENT APPLICATION NUMBER: US/09/929,955
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/225,767
; PRIOR FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-09-929-955-4

Query Match      50.0%; Score 10; DB 10; Length 350;
Best Local Similarity 100.0%; Pred. No. 0.0076;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
    |||||
Db 30 QNIQLINTNG 39

RESULT 9
US-10-104-966-4
; Sequence 4, Application US/10104966
; Publication No. US20020155124A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.23AUSC1
; CURRENT APPLICATION NUMBER: US/10/104,966
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-10-104-966-4

Query Match      50.0%; Score 10; DB 14; Length 350;
Best Local Similarity 100.0%; Pred. No. 0.0076;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
    |||||
Db 30 QNIQLINTNG 39
```

```
RESULT 10
US-10-128-587A-97
; Sequence 97, Application US/10128587A
; Publication No. US20030152940A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Constructs and methods for expression of recombinant HCV envelope
; FILE REFERENCE: 134 PCT
; CURRENT FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: hepatitis C virus
US-10-128-587A-97

Query Match      50.0%; Score 10; DB 12; Length 363;
Best Local Similarity 100.0%; Pred. No. 0.0079;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QNIQLINTNG 10
        |||||
Db      26 QNIQLINTNG 35

RESULT 11
US-10-128-590-97
; Sequence 97, Application US/10128590
; Publication No. US20030108561A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Core-glycosylated HCV envelope proteins
; FILE REFERENCE: 135 PCT
; CURRENT FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-128-590-97

Query Match      50.0%; Score 10; DB 15; Length 363;
Best Local Similarity 100.0%; Pred. No. 0.0079;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QNIQLINTNG 10
        |||||
Db      26 QNIQLINTNG 35

RESULT 12
US-09-941-611-23
; Sequence 23, Application US/09941611
; Patent No. US20020106640A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; POLLET, DIRK
; MAERTENS, GEERT
; VAN HEUVERSWUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; ANTI-BODIES TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
```

```
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/941.611
FILING DATE: 30-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: 1995-02-21
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-941-611-23

Query Match      50.0%; Score 10; DB 10; Length 2894;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QNIQLINTNG 10
        |||||
Db      409 QNIQLINTNG 418

RESULT 13
US-10-044-995-23
; Sequence 23, Application US/10044995
; Publication No. US20030049685A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; POLLET, DIRK
; MAERTENS, GEERT
; VAN HEUVERSWUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; ANTI-BODIES TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
```

CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/044,995
; FILING DATE: 15-Jan-2002
; CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/391,671
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 07/920,286
; FILING DATE: 14-OCT-1992
; APPLICATION NUMBER: WO PCT/EP91/02409
; FILING DATE: 13-DEC-1991
; APPLICATION NUMBER: EP 90124241.2
; FILING DATE: 14-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
; TELEPHONE: 7038164000
; TELEFAX: 7038164100
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-044-995-23

Query Match 50.0%; Score 10; DB 15; Length 2894;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
|||||

Db 409 QNIQLINTNG 418

RESULT 14
US-09-742-659-4
; Sequence 4, Application US/09742659
; Patent No. US20010034019A1
; GENERAL INFORMATION:
; APPLICANT: Hong, Zhi
; APPLICANT: Butkiewicz, Nancy J.
; APPLICANT: Zhong, Weidong
; APPLICANT: Ingravallo, Paul
; APPLICANT: Wright-Minogue, Jacquelyn
; APPLICANT: Lau, Johnson Y.
; APPLICANT: Lemon, Stanley M.
; TITLE OF INVENTION: Chimeric HCV/GBV-B viruses
; FILE REFERENCE: ID0116
; CURRENT APPLICATION NUMBER: US/09/742,659
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: US 60/171,469
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-742-659-4

Query Match 50.0%; Score 10; DB 9; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.045;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
|||||

Db 409 QNIQLINTNG 418

RESULT 15

US-09-238-076-20
; Sequence 20, Application US/09238076
; Patent No. US20020102540A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; TITLE OF INVENTION: VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,076
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/034,756
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
US-09-238-076-20

Query Match 50.0%; Score 10; DB 10; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.045;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
|||||

Db 409 QNIQLINTNG 418

Search completed: November 21, 2003, 22:19:36
Job time : 19.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 163.1 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-73

Perfect score: 20

Sequence: 1 QNQLINTNGQWHINSTALN 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents AA Main:*

- 1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/1/paa/US097A_COMB.pep.*
- 22: /cgn2_6/ptodata/1/paa/US097B_COMB.pep.*
- 23: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
- 24: /cgn2_6/ptodata/1/paa/US099A_COMB.pep.*
- 25: /cgn2_6/ptodata/1/paa/US099B_COMB.pep.*
- 26: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
- 27: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
- 28: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
- 29: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
- 30: /cgn2_6/ptodata/1/paa/US104_COMB.pep.*
- 31: /cgn2_6/ptodata/1/paa/US106_COMB.pep.*
- 32: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	20	100.0	20	23	US-09-899-303-73

2	20	100.0	20	23	US-09-899-303A-73	Sequence 73, Appl
3	20	100.0	20	25	US-09-973-025-73	Sequence 73, Appl
4	20	100.0	20	25	US-09-995-791-73	Sequence 73, Appl
5	20	100.0	20	25	US-09-995-808-73	Sequence 73, Appl
6	20	100.0	20	25	US-09-995-860-73	Sequence 73, Appl
7	20	100.0	20	26	US-10-020-510-73	Sequence 73, Appl
8	20	100.0	20	29	US-10-321-798-73	Sequence 73, Appl
9	10	50.0	20	13	US-08-974-685-47	Sequence 47, Appl
10	10	50.0	20	13	US-08-974-685-47	Sequence 47, Appl
11	10	50.0	20	13	US-08-974-690A-47	Sequence 47, Appl
12	10	50.0	20	13	US-08-974-690B-47	Sequence 47, Appl
13	10	50.0	20	13	US-08-974-690C-47	Sequence 47, Appl
14	10	50.0	52	8	US-08-475-482-94	Sequence 94, Appl
15	10	50.0	52	8	US-08-477-072-94	Sequence 94, Appl
16	10	50.0	52	8	US-08-477-582-94	Sequence 94, Appl
17	10	50.0	52	8	US-08-480-253-94	Sequence 94, Appl
18	10	50.0	57	9	US-08-507-740-17	Sequence 17, Appl
19	10	50.0	57	13	US-08-505-054-17	Sequence 17, Appl
20	10	50.0	57	13	US-08-505-054B-17	Sequence 17, Appl
21	10	50.0	91	24	US-09-921-397-80	Sequence 80, Appl
22	10	50.0	139	3	US-07-722-489-520	Sequence 520, Appl
23	10	50.0	139	3	US-07-722-489-523	Sequence 523, Appl
24	10	50.0	139	8	US-08-403-590B-171	Sequence 171, Appl
25	10	50.0	139	8	US-08-403-590B-174	Sequence 174, Appl
26	10	50.0	139	8	US-08-444-112-171	Sequence 171, Appl
27	10	50.0	139	8	US-08-444-112-174	Sequence 174, Appl
28	10	50.0	144	8	US-08-471-498-20	Sequence 20, Appl
29	10	50.0	160	28	US-10-292-129-15	Sequence 15, Appl
30	10	50.0	278	8	US-08-471-498-15	Sequence 15, Appl
31	10	50.0	333	8	US-08-417-478-12	Sequence 12, Appl
32	10	50.0	333	8	US-08-453-613-12	Sequence 12, Appl
33	10	50.0	339	3	US-07-722-489-499	Sequence 499, Appl
34	10	50.0	339	3	US-07-722-489-507	Sequence 507, Appl
35	10	50.0	339	8	US-08-403-590B-150	Sequence 150, Appl
36	10	50.0	339	8	US-08-403-590B-158	Sequence 158, Appl
37	10	50.0	339	8	US-08-444-112-150	Sequence 150, Appl
38	10	50.0	339	8	US-08-444-112-158	Sequence 158, Appl
39	10	50.0	350	21	US-09-705-547-4	Sequence 4, Appl
40	10	50.0	350	24	US-09-929-955-4	Sequence 4, Appl
41	10	50.0	350	27	US-10-104-966-4	Sequence 4, Appl
42	10	50.0	350	22	US-60-229-175-4	Sequence 4, Appl
43	10	50.0	363	27	US-10-128-587A-97	Sequence 97, Appl
44	10	50.0	363	27	US-10-128-590-97	Sequence 97, Appl
45	10	50.0	367	8	US-08-417-478-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1

US-09-899-303-73

; Sequence 73, Application US/09899303

; GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

BOSMAN, FONS

DE MARTYNOFF, GUY

BUYSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/899,303
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-09-899-303-73

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNGQWHINSTALN 20
Db 1 QNIQLINTNGQWHINSTALN 20

RESULT 2

US-09-899-303A-73
Sequence 73, Application US/09899303A

GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303A
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/612,973
FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-09-899-303A-73

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNGQWHINSTALN 20
Db 1 QNIQLINTNGQWHINSTALN 20

RESULT 3

US-09-973-025-73
Sequence 73, Application US/09973025

GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/973,025
FILING DATE: 10-Oct-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-09-973-025-73

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNGQWHINSTALN 20
Db 1 QNIQLINTNGQWHINSTALN 20

RESULT 4

US-09-995-791-73

; Sequence 73, Application US/09995791
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 73
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-73

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNGQWHINSTALN 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 QNIQLINTNGQWHINSTALN 20

RESULT 5
US-09-995-808-73
; Sequence 73, Application US/09995808
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 73
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-73

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNGQWHINSTALN 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 QNIQLINTNGQWHINSTALN 20

RESULT 6
US-09-995-860-73
; Sequence 73, Application US/09995860
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 73
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-73

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNGQWHINSTALN 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 QNIQLINTNGQWHINSTALN 20

RESULT 7
US-10-020-510-73
; Sequence 73, Application US/10020510
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-72
; CURRENT APPLICATION NUMBER: US/10/020,510
; CURRENT FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 73
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-020-510-73

Query Match 100.0%; Score 20; DB 26; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNGQWHINSTALN 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 QNIQLINTNGQWHINSTALN 20

RESULT 8
US-10-321-798-73
; Sequence 73, Application US/10321798
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-93
; CURRENT APPLICATION NUMBER: US/10/321,798
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: 60/418,358
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: 10/020,510
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 73
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-321-798-73

Query Match 100.0%; Score 20; DB 29; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNGQWHINSTALN 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 QNIQLINTNGQWHINSTALN 20

RESULT 9
US-08-974-685-47
; Sequence 47, Application US/08974685
; GENERAL INFORMATION:
; APPLICANT: LEROUIX-ROELS, GEERT
; DELEYS, ROBERT
; MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 181
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,685
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-08-974-685-47

Query Match 50.0%; Score 10; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
Db 1 QNIQLINTNG 10

RESULT 10

US-08-974-690-47
Sequence 47, Application US/08974690
GENERAL INFORMATION:
APPLICANT: LEROUIX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
APPLICANT: MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 166
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/635,886

FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-974-690-47

Query Match 50.0%; Score 10; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
Db 1 QNIQLINTNG 10

RESULT 11

US-08-974-690A-47
Sequence 47, Application US/08974690A
GENERAL INFORMATION:
APPLICANT: LEROUIX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
APPLICANT: MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690A
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-08-974-690A-47

Query Match 50.0%; Score 10; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNIQLINTNG 10
| | | | |
Db 1 QNIQLINTNG 10

RESULT 12
US-08-974-690B-47
; Sequence 47, Application US/08974690B
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; DELEYS, ROBERT
; MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
; THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA: US/08/974,690B
; APPLICATION NUMBER: US/08/974,690B
; FILING DATE: 19-Nov-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B. J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 2752-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-08-974-690B-47

Query Match 50.0%; Score 10; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNIQLINTNG 10
| | | | |
Db 1 QNIQLINTNG 10

RESULT 13
US-08-974-690C-47
; Sequence 47, Application US/08974690C
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; DELEYS, ROBERT
; MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/08/974,690C
; CURRENT FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28

; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 47
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-974-690C-47

Query Match 50.0%; Score 10; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNIQLINTNG 10
| | | | |
Db 1 QNIQLINTNG 10

RESULT 14
US-08-475-482-94
; Sequence 94, Application US/08475482
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA: US/08/475,482
; APPLICATION NUMBER: US/08/475,482
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 17-June-1994
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-475-482-94

Query Match 50.0%; Score 10; DB 8; Length 52;
Best Local Similarity 100.0%; Pred. No. 0.0042;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNIQLINTNG 10
| | | | |
DB 2 QNIQLINTNG 11

RESULT 15

US-08-477-072-94
; Sequence 94, Application US/08477072
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,072
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 17-June-1994
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-477-072-94

Query Match 50.0%; Score 10; DB 8; Length 52;
Best Local Similarity 100.0%; Pred. No. 0.0042;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNIQLINTNG 10
| | | | |
DB 2 QNIQLINTNG 11

Search completed: November 21, 2003, 22:09:49

Job time : 163.1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.55 Seconds
(without alignments)
117.536 Million cell updates/sec

Title: US-09-973-025-73

Perfect score: 20

Sequence: 1 QNIQLINTNGQWHINSTALN 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents AA New:*

- 1: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep:*
- 2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep:*
- 3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep:*
- 4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep:*
- 5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:*
- 6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:*
- 7: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	50.0	20	6	US-10-651-165-47
2	10	50.0	30	6	US-10-296-734-460
3	10	50.0	30	6	US-10-296-734-462
4	10	50.0	450	6	US-10-651-165-179
5	10	50.0	450	6	US-10-651-165-181
6	10	50.0	1997	6	US-10-296-734-816
7	10	50.0	2010	6	US-10-296-734-814
8	10	50.0	3011	6	US-10-296-734-406
9	10	50.0	5985	6	US-10-296-734-810
10	9	45.0	30	6	US-10-296-734-464
11	8	40.0	30	6	US-10-651-165-116
12	8	40.0	9	6	US-10-651-165-115
13	8	40.0	20	6	US-10-651-165-46
14	8	40.0	450	6	US-10-651-165-190
15	8	40.0	450	6	US-10-651-165-191
16	8	40.0	450	6	US-10-651-165-192
17	8	40.0	450	6	US-10-651-165-193
18	8	40.0	450	6	US-10-651-165-195
19	8	40.0	450	6	US-10-651-165-200
20	8	40.0	1026	1	PCT-US03-20409-3
21	8	40.0	2280	1	PCT-US03-20322-211
22	7	35.0	9	6	US-10-651-165-113
23	7	35.0	176	5	US-09-638-693-158
24	7	35.0	176	5	US-09-873-224A-158
25	7	35.0	333	1	PCT-US03-19834-3
26	7	35.0	333	6	US-10-655-562-4

Sequence 180, Appl
Sequence 4, Appl
Sequence 2, Appl
Sequence 114, Appl
Sequence 14, Appl
Sequence 16, Appl
Sequence 19, Appl
Sequence 21, Appl
Sequence 29, Appl
Sequence 30, Appl
Sequence 44, Appl
Sequence 51, Appl
Sequence 52, Appl
Sequence 53, Appl
Sequence 54, Appl
Sequence 57, Appl
Sequence 58, Appl
Sequence 61, Appl
Sequence 62, Appl

27 7 35.0 450 6 US-10-651-165-180
28 7 35.0 637 1 PCT-US03-33610-4
29 7 35.0 3011 1 PCT-US03-19834-2
30 6 30.0 8 6 US-10-651-165-114
31 6 30.0 31 4 US-08-823-980F-14
32 6 30.0 31 4 US-08-823-980F-16
33 6 30.0 31 4 US-08-823-980F-19
34 6 30.0 31 4 US-08-823-980F-21
35 6 30.0 31 4 US-08-823-980F-29
36 6 30.0 31 4 US-08-823-980F-30
37 6 30.0 31 4 US-08-823-980F-44
38 6 30.0 31 4 US-08-823-980F-51
39 6 30.0 31 4 US-08-823-980F-52
40 6 30.0 31 4 US-08-823-980F-53
41 6 30.0 31 4 US-08-823-980F-54
42 6 30.0 31 4 US-08-823-980F-57
43 6 30.0 31 4 US-08-823-980F-58
44 6 30.0 31 4 US-08-823-980F-61
45 6 30.0 31 4 US-08-823-980F-62

ALIGNMENTS

RESULT 1

US-10-651-165-47
; Sequence 47, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-47

Query Match 50.0%; Score 10; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.00033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
| | | | | | | | | |

Db 1 QNIQLINTNG 10
| | | | | | | | | |

RESULT 2

US-10-296-734-460
; Sequence 460, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 460
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 27
US-10-296-734-460

Query Match 50.0%; Score 10; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.00045;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
| | | | | | | |
Db 21 QNIQLINTNG 30

RESULT 3
US-10-296-734-462
; Sequence 462, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; PRIOR FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 462
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 28
US-10-296-734-462

Query Match 50.0%; Score 10; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.00045;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
| | | | | | | |
Db 6 QNIQLINTNG 15

RESULT 4
US-10-651-165-179
; Sequence 179, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 179
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-179

Query Match 50.0%; Score 10; DB 6; Length 450;
Best Local Similarity 100.0%; Pred. No. 0.0036;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
| | | | | | | |
Db 409 QNIQLINTNG 418

RESULT 5
US-10-651-165-181
; Sequence 181, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 181
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-181

Query Match 50.0%; Score 10; DB 6; Length 450;
Best Local Similarity 100.0%; Pred. No. 0.0036;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
| | | | | | | |
Db 409 QNIQLINTNG 418

RESULT 6
US-10-296-734-816
; Sequence 816, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 816
; LENGTH: 1997
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette C
US-10-296-734-816

Query Match 50.0%; Score 10; DB 6; Length 1997;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
| | | | | | | |

```
Db      576 QNIQLINTNG 585

RESULT 7
US-10-296-734-814
; Sequence 814, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 814
; LENGTH: 2010
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette B
US-10-296-734-814

Query Match      50.0%; Score 10; DB 6; Length 2010;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QNIQLINTNG 10
      |||||
Db      171 QNIQLINTNG 180

RESULT 8
US-10-296-734-406
; Sequence 406, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la consensus polyprotein
US-10-296-734-406

Query Match      50.0%; Score 10; DB 6; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QNIQLINTNG 10
      |||||
Db      409 QNIQLINTNG 418

RESULT 9
US-10-296-734-810
; Sequence 810, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
```

```
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 810
; LENGTH: 5985
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la savine
US-10-296-734-810

Query Match      50.0%; Score 10; DB 6; Length 5985;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QNIQLINTNG 10
      |||||
Db      2181 QNIQLINTNG 2190

RESULT 10
US-10-296-734-464
; Sequence 464, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 464
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la segment 29
US-10-296-734-464

Query Match      45.0%; Score 9; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.0046;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      12 WHINSTALN 20
      |||||
Db      2 WHINSTALN 10

RESULT 11
US-10-651-165-116
; Sequence 116, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 116
; LENGTH: 8
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-116

Query Match      40.0%; Score 8; DB 6; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.1e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      12 WHINSTAL 19
Db      1 WHINSTAL 8

RESULT 12
US-10-651-165-115
; Sequence 115, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651.165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 115
; LENGTH: 9
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-115

Query Match      40.0%; Score 8; DB 6; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.1e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      12 WHINSTAL 19
Db      2 WHINSTAL 9

RESULT 13
US-10-651-165-46
; Sequence 46, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651.165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-190
```

```
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-46

Query Match      40.0%; Score 8; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QNIQILINT 8
Db      13 QNIQILINT 20

RESULT 14
US-10-651-165-190
; Sequence 190, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651.165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 190
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-190

Query Match      40.0%; Score 8; DB 6; Length 450;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      3 IQILINTNG 10
Db      411 IQILINTNG 418

RESULT 15
US-10-651-165-191
; Sequence 191, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651.165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 191
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-191
```


Query Match 40.0%; Score 8; DB 6; Length 450;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 IQLINTG 10
|
|
|
|
|
|
|
|
Db 411 IQLINTG 418

Search completed: November 21, 2003, 22:12:53
Job time : 9.55 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:46:27 ; Search time 9.25 Seconds
(without alignments)
207.932 Million cell updates/sec

Title: US-09-973-025-73

Perfect score: 20

Sequence: 1 QNIQLINTNGQWHINSTALN 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	50.0	101	2 PS0449	genome polyprotein
2	10	50.0	133	2 D48776	polyprotein (E2/NS
3	10	50.0	135	2 S24086	envelope protein -
4	10	50.0	135	2 PS0448	genome polyprotein
5	10	50.0	137	2 S24088	envelope protein -
6	10	50.0	138	2 S24100	envelope protein -
7	10	50.0	138	2 PC1205	envelope protein -
8	10	50.0	138	2 PC1203	envelope protein -
9	10	50.0	138	2 PC1206	envelope protein -
10	10	50.0	138	2 PC1208	envelope protein -
11	10	50.0	138	2 PC1209	envelope protein -
12	10	50.0	138	2 PC1210	envelope protein -
13	10	50.0	138	2 PC1211	envelope protein -
14	10	50.0	138	2 PC1212	envelope protein -
15	10	50.0	138	2 PC1197	genome polyprotein
16	10	50.0	138	2 PC1200	genome polyprotein
17	10	50.0	138	2 PC1201	envelope protein -
18	10	50.0	138	2 PC1204	envelope protein -
19	10	50.0	138	2 S24107	envelope protein -
20	10	50.0	138	2 S24103	envelope protein -
21	10	50.0	350	2 S35631	genome polyprotein
22	10	50.0	513	2 A44150	structural protein
23	10	50.0	640	2 JQ1584	genome polyprotein
24	10	50.0	3011	1 GNMVCH	genome polyprotein
25	10	50.0	3011	1 S40770	genome polyprotein
26	9	45.0	138	2 PC1207	envelope protein -
27	9	45.0	138	2 S24070	envelope protein -
28	9	45.0	138	2 S24074	envelope protein -
29	9	45.0	138	2 S24106	envelope protein -

30 9 45.0 141 2 S24098 envelope protein -
31 9 45.0 492 2 S41288 genome polyprotein
32 9 45.0 3014 1 JC5620 genome polyprotein
33 8 40.0 127 2 B48776 polyprotein (E2/NS
34 8 40.0 134 2 S24089 envelope protein -
35 8 40.0 136 2 S24090 envelope protein -
36 8 40.0 136 2 S24091 genome polyprotein
37 8 40.0 138 2 PC1199 genome polyprotein
38 8 40.0 138 2 S24102 envelope protein -
39 8 40.0 138 2 S24099 envelope protein -
40 8 40.0 138 2 S24083 envelope protein -
41 8 40.0 138 2 S24079 envelope protein -
42 8 40.0 138 2 S24104 envelope protein -
43 8 40.0 138 2 S24101 envelope protein -
44 8 40.0 138 2 S24068 envelope protein -
45 8 40.0 138 2 S24069 envelope protein -

ALIGNMENTS

RESULT 1

PS0449 genome polyprotein - hepatitis C virus (strain PRCl1) (fragments)
N:Contains: envelope protein E2; nonstructural protein NS1; nonstructural protein NS5
C:Species: hepatitis C virus
C>Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
C:Accession: PS0449

R:Liu, K.; Hu, Z.; Li, H.; Prince, A.M.; Inchauspe, G.

Gene 114, 245-250, 1992

A:Title: Genomic typing of hepatitis C viruses present in China.

A:Reference number: JH0711; MUID:92290283; PMID:1318245

A:Accession: PS0449

A>Status: nucleic acid sequence not shown

A:Molecule type: genomic RNA

A:Residues: 1-101 <Liu>

A:Cross-references: GB:M74887; GB:M74892

A>Note: translation of the nucleotide sequence is not complete

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: envelope protein; glycoprotein; nonstructural protein; polyprotein

F:1-65/Product: envelope protein E2 or nonstructural protein NS1 (fragment) #status predicted

F:66-101/Product: nonstructural protein NS5 (fragment) #status predicted <NS5>

F:37,43,50/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 50.0%; Score 10; DB 2; Length 101;
Best Local Similarity 100.0%; Pred. No. 0.00053;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10

Db 29 QNIQLINTNG 38

RESULT 2

D48776 polyprotein (E2/NS1 region, HVR1, HVR2) - hepatitis C virus (fragment)

C:Species: hepatitis C virus

C>Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 17-Nov-2000

C:Accession: D48776

R:Higashi, Y.; Kakumu, S.; Yoshioka, K.; Wakita, T.; Mizokami, M.; Ohba, K.; Ito, Y.; Is

Virolgy 197, 659-668, 1993

A:Title: Dynamics of genome change in the E2/NS1 region of hepatitis C virus in vivo.

A:Reference number: A48776; MUID:94069940; PMID:8249288

A:Accession: D48776

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-133 <HIG>

A:Experimental source: subtype III, patient KS

A>Note: sequence extracted from NCBI backbone (NCBI:140217)

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: polyprotein

Query Match 50.0%; Score 10; DB 2; Length 133;

Best Local Similarity 100.0%; Pred. No. 0.00069;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
|||||
Db 38 QNIQLINTNG 47

RESULT 3
S24086
envelope protein - hepatitis C virus (fragment)
N;Contains: envelope protein gp35; envelope protein gp70
C;Species: hepatitis C virus
C;Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C;Accession: S24086; S78228
R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraio, K.; Ohkoshi,
Virus Res. 22, 107-123, 1992
A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A;Reference number: S24068; MUID:92230394; PMID:1314471
A;Accession: S24086
A;Molecule type: genomic RNA
A;Residues: 1-135 <KAT>
A;Cross-references: EMBL:X60572
A;Experimental source: isolate RE55A
R;Kato, N.
submitted to the EMBL Data Library, August 1991
A;Reference number: S78226
A;Accession: S78228
A;Molecule type: genomic RNA
A;Residues: 1-15, 'XXX', 16-135 <KAW>
A;Cross-references: EMBL:X60572
C;Superfamily: hepatitis C virus genome polypeptide
C;Keywords: envelope protein; glycoprotein; polypeptide
F;1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F;16-135/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F;16-39/Region: hypervariable 1 #status predicted
F;103-109/Region: hypervariable 2 #status predicted
F;46,52,59,77/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 50.0%; Score 10; DB 2; Length 135;
Best Local Similarity 100.0%; Pred. No. 0.0007;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
|||||
Db 38 QNIQLINTNG 47

RESULT 4
PS0448
genome polypeptide - hepatitis C virus (strain PR4) (fragments)
N;Contains: envelope protein E1; envelope protein E2; nonstructural protein NS1; nonstru
C;Species: hepatitis C virus
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
C;Accession: PS0448
R;Liu, K.; Hu, Z.; Li, H.; Prince, A.M.; Inchauspe, G.
Gene 114, 245-250, 1992
A;Title: Genomic typing of hepatitis C viruses present in China.
A;Reference number: JH0711; MUID:92290283; PMID:1318245
A;Accession: PS0448
A;Status: nucleic acid sequence not shown
A;Molecule type: genomic RNA
A;Residues: 1-135 <LIU>
A;Cross-references: GB:M74886
A;Note: translation of the nucleotide sequence is not complete
C;Superfamily: hepatitis C virus genome polypeptide
C;Keywords: envelope protein; glycoprotein; nonstructural protein; polypeptide
F;1-10/Product: envelope protein E1 (fragment) #status predicted <EPI>
F;11-99/Product: envelope protein E2 or nonstructural protein NS1 (fragment) #status pre
F;100-135/Product: nonstructural protein NS5 (fragment) #status predicted <NS>
F;47,53,60,78/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 50.0%; Score 10; DB 2; Length 135;

Best Local Similarity 100.0%; Pred. No. 0.0007;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
|||||
Db 39 QNIQLINTNG 48

RESULT 5
S24088
envelope protein - hepatitis C virus (fragment)
N;Contains: envelope protein gp35; envelope protein gp70
C;Species: hepatitis C virus
C;Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C;Accession: S24088; S78229
R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraio, K.; Ohkoshi,
Virus Res. 22, 107-123, 1992
A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A;Reference number: S24068; MUID:92230394; PMID:1314471
A;Accession: S24088
A;Molecule type: genomic RNA
A;Residues: 1-137 <KAT>
A;Cross-references: EMBL:X60574
A;Experimental source: isolate RE55C
R;Kato, N.
submitted to the EMBL Data Library, August 1991
A;Reference number: S78226
A;Accession: S78229
A;Molecule type: genomic RNA
A;Residues: 1-15, 'X', 16-137 <KAW>
A;Cross-references: EMBL:X60574
C;Superfamily: hepatitis C virus genome polypeptide
C;Keywords: envelope protein; glycoprotein; polypeptide
F;1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F;16-137/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F;16-41/Region: hypervariable 1 #status predicted
F;105-111/Region: hypervariable 2 #status predicted
F;48,54,61,79/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 50.0%; Score 10; DB 2; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
|||||
Db 40 QNIQLINTNG 49

RESULT 6
S24100
envelope protein - hepatitis C virus (fragment)
N;Contains: envelope protein gp35; envelope protein gp70
C;Species: hepatitis C virus
C;Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C;Accession: S24100; S78234
R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraio, K.; Ohkoshi,
Virus Res. 22, 107-123, 1992
A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A;Reference number: S24068; MUID:92230394; PMID:1314471
A;Accession: S24100
A;Molecule type: genomic RNA
A;Residues: 1-138 <KAT>
A;Cross-references: EMBL:X60583
A;Experimental source: isolate RE71
R;Kato, N.
submitted to the EMBL Data Library, August 1991
A;Reference number: S78226
A;Accession: S78234
A;Molecule type: genomic RNA
A;Residues: 1-83, 'C', 85-138 <KAW>
A;Cross-references: EMBL:X60583
C;Superfamily: hepatitis C virus genome polypeptide
C;Keywords: envelope protein; glycoprotein; polypeptide

F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted
F:49,55,62,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 50.0%; Score 10; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
|||||
Db 41 QNIQLINTNG 50

RESULT 7

PC1205
envelope protein - hepatitis C virus (strain RS3-2) (fragment)

C:Species: hepatitis C virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C:Accession: PC1205
R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Sekiya, H.; Hijikata, M.; Shimotoh
Biochem. Biophys. Res. Commun. 189, 119-127, 1992
A:Title: Characterization of hypervariable regions in the putative envelope protein of h
A:Reference number: PC1182; MUID:93080545; PMID:1333186

A:Accession: PC1205
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-138 <KAT>
A:Cross-references: DBJ:D12942; DDBJ:D12972
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein

Query Match 50.0%; Score 10; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
|||||
Db 41 QNIQLINTNG 50

RESULT 8

PC1203

envelope protein - hepatitis C virus (strain RS2-4) (fragment)

C:Species: hepatitis C virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C:Accession: PC1203
R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Sekiya, H.; Hijikata, M.; Shimotoh
Biochem. Biophys. Res. Commun. 189, 119-127, 1992
A:Title: Characterization of hypervariable regions in the putative envelope protein of h
A:Reference number: PC1182; MUID:93080545; PMID:1333186

A:Accession: PC1203
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-138 <KAT>
A:Cross-references: DBJ:D12942; DDBJ:D12972
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein

Query Match 50.0%; Score 10; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
|||||
Db 41 QNIQLINTNG 50

RESULT 9

PC1206

envelope protein - hepatitis C virus (strain RS3-3) (fragment)

C:Species: hepatitis C virus

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C:Accession: PC1206
R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Sekiya, H.; Hijikata, M.; Shimotoh
Biochem. Biophys. Res. Commun. 189, 119-127, 1992
A:Title: Characterization of hypervariable regions in the putative envelope protein of h
A:Reference number: PC1182; MUID:93080545; PMID:1333186

A:Accession: PC1206
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-138 <KAT>
A:Cross-references: DDBJ:D12942; DDBJ:D12972
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein

Query Match 50.0%; Score 10; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
|||||
Db 41 QNIQLINTNG 50

RESULT 10

PC1208

envelope protein - hepatitis C virus (strain RS4-2) (fragment)

C:Species: hepatitis C virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C:Accession: PC1208
R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Sekiya, H.; Hijikata, M.; Shimotoh
Biochem. Biophys. Res. Commun. 189, 119-127, 1992
A:Title: Characterization of hypervariable regions in the putative envelope protein of h
A:Reference number: PC1182; MUID:93080545; PMID:1333186

A:Accession: PC1208
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-138 <KAT>
A:Cross-references: DDBJ:D12942; DDBJ:D12972
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein

Query Match 50.0%; Score 10; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
|||||
Db 41 QNIQLINTNG 50

RESULT 11

PC1209

envelope protein - hepatitis C virus (strain RS4-3) (fragment)

C:Species: hepatitis C virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C:Accession: PC1209
R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Sekiya, H.; Hijikata, M.; Shimotoh
Biochem. Biophys. Res. Commun. 189, 119-127, 1992
A:Title: Characterization of hypervariable regions in the putative envelope protein of h
A:Reference number: PC1182; MUID:93080545; PMID:1333186

A:Accession: PC1209
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-138 <KAT>
A:Cross-references: DDBJ:D12942; DDBJ:D12972
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein

Query Match 50.0%; Score 10; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10

```
Db          41 QNIQLINTNG 50

RESULT 12
PC1210
envelope protein - hepatitis C virus (strain RS5-1) (fragment)
C;Species: hepatitis C virus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C;Accession: PC1210
R;Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Sekiya, H.; Hijikata, M.; Shimotoh
Biochem. Biophys. Res. Commun. 189, 119-127, 1992
A;Title: Characterization of hypervariable regions in the putative envelope protein of h
A;Reference number: PC1182; MUID:93080545; PMID:1333186
A;Accession: PC1210
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-138 <KAT>
A;Cross-references: DBJ:D12942; DBJ:D12972
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: envelope protein

Query Match      50.0%; Score 10; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QNIQLINTNG 10
      |||||
Db      41 QNIQLINTNG 50

RESULT 13
PC1211
envelope protein - hepatitis C virus (strain RS5-2) (fragment)
C;Species: hepatitis C virus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C;Accession: PC1211
R;Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Sekiya, H.; Hijikata, M.; Shimotoh
Biochem. Biophys. Res. Commun. 189, 119-127, 1992
A;Title: Characterization of hypervariable regions in the putative envelope protein of h
A;Reference number: PC1182; MUID:93080545; PMID:1333186
A;Accession: PC1211
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-138 <KAT>
A;Cross-references: DBJ:D12942; DBJ:D12972
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: envelope protein

Query Match      50.0%; Score 10; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QNIQLINTNG 10
      |||||
Db      41 QNIQLINTNG 50

RESULT 14
PC1212
envelope protein - hepatitis C virus (strain RS5-3) (fragment)
C;Species: hepatitis C virus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C;Accession: PC1212
R;Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Sekiya, H.; Hijikata, M.; Shimotoh
Biochem. Biophys. Res. Commun. 189, 119-127, 1992
A;Title: Characterization of hypervariable regions in the putative envelope protein of h
A;Reference number: PC1182; MUID:93080545; PMID:1333186
A;Accession: PC1212
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-138 <KAT>
A;Cross-references: DBJ:D12942; DBJ:D12972; NID:g285884; PIDN:BAA02347.1; PID:g285885
```

C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: envelope protein

Query Match 50.0%; Score 10; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
 |||||
Db 41 QNIQLINTNG 50

RESULT 15

PC1197

genome polyprotein - hepatitis C virus (strain RS1-1) (fragment)

C;Species: hepatitis C virus

C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000

C;Accession: PC1197

R;Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Sekiya, H.; Hijikata, M.; Shimotoh

Biochem. Biophys. Res. Commun. 189, 119-127, 1992

A;Title: Characterization of hypervariable regions in the putative envelope protein of h

A;Reference number: PC1182; MUID:93080545; PMID:1333186

A;Accession: PC1197

A;Status: nucleic acid sequence not shown

A;Molecule type: genomic RNA

A;Residues: 1-138 <KAT>

A;Cross-references: GB:D12957; NID:g285854; PIDN:BAA02333.1; PID:g285855

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: polyprotein

Query Match 50.0%; Score 10; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
 |||||
Db 41 QNIQLINTNG 50

Search completed: November 21, 2003, 21:11:29

Job time : 10.25 sec

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:52 ; Search time 4.85 seconds
(without alignments)
193.925 Million cell updates/sec

Title: US-09-973-025-73
Perfect score: 20
Sequence: 1 QN1QIINTNGQWHINSTALN 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	10	50.0	138	1 POLG_HCVE0
2	10	50.0	321	1 POLG_HCVTH
3	10	50.0	3011	1 POLG_HCVH
4	9	45.0	321	1 POLG_HCVH8
5	8	40.0	309	1 POLG_HCVH7
6	8	40.0	513	1 POLG_HCVJ2
7	8	40.0	520	1 POLG_HCVH4
8	8	40.0	520	1 POLG_HCVH8
9	8	40.0	3010	1 POLG_HCVBK
10	8	40.0	3010	1 POLG_HCVJT
11	8	40.0	3010	1 POLG_HCVTW
12	8	40.0	3033	1 POLG_HCVJ6
13	7	35.0	526	1 PUR9_BUCAP
14	7	35.0	737	1 POLG_HCVU5
15	7	35.0	3011	1 POLG_HCV1
16	6	30.0	146	1 YN59_YEAST
17	6	30.0	158	1 YBKA_HAEIN
18	6	30.0	258	1 YCU9_YEAST
19	6	30.0	552	1 PRG_WIGER
20	6	30.0	577	1 SYR_HAEIN
21	6	30.0	601	1 LEPA_BUCAP
22	6	30.0	654	1 RPK3_YEAST
23	6	30.0	659	1 DNLJ_MYCBE
24	6	30.0	737	1 POLG_HCVU7
25	6	30.0	747	1 PARC_HAEIN
26	6	30.0	910	1 HUL5_YEAST
27	6	30.0	1436	1 WIN1_SCHPO
28	6	30.0	3033	1 POLG_HCVJ8
29	5	25.0	43	1 CPC6_CANPG
30	5	25.0	112	1 OLI2_MOUSE
31	5	25.0	116	1 CN4D_MOUSE
32	5	25.0	129	1 WNT6_XENLA
33	5	25.0	135	1 SPAK_SALTY

34 5 25.0 143 1 USPA_ECOLI
35 5 25.0 149 1 NPC2_MOUSE
36 5 25.0 150 1 RL9_LACLA
37 5 25.0 158 1 NUD8_HAEIN
38 5 25.0 163 1 Y207_MYCGE
39 5 25.0 166 1 RL21_ENTHI
40 5 25.0 167 1 FIMG_ECOLI
41 5 25.0 179 1 YC21_ANTSP
42 5 25.0 194 1 LFCA_HAEPU
43 5 25.0 202 1 FM3_KLEPN
44 5 25.0 203 1 FCP_ODOSI
45 5 25.0 205 1 ENGB_HAEIN

ALIGNMENTS

RESULT 1
POLG_HCVE0
ID POLG_HCVE0 STANDARD; PRT; 138 AA.
AC P27953;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1)] (Fragment).
OS Hepatitis C virus (isolate EC10) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11106;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J., Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M., Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins.";
RL Virology 180:842-848(1991).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; X53136; CAA37296.1; -
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
Transmembrane; Nonstructural protein.
FT NON_TER 1 84
FT CHAIN 6
FT CHAIN 85 >138
FT CARBOHYD 6
FT CARBOHYD 118
FT CARBOHYD 124 124
FT CARBOHYD 131 131
FT NON_TER 138 138
SQ SEQUENCE 138 AA; 14781 MW; CD3F0A962DEAB1AD CRC64;
Query Match 50.0%; Score 10; DB 1; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.00039;

Matches	10;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	----	--------	----	------	----

Qy 1 QNIQLNTG 10
Db 110 QNIQLNTG 119

```

RESULT 2
POLG_HCVTH      STANDARD;      PRRT;      321 AA.
ID              AC              P27957;
DT              01-AUG-1992 (Rel. 23, Created)
DT              01-AUG-1992 (Rel. 23, Last sequence update)
DT              16-OCT-2001 (Rel. 40, Last annotation update)
DE              Genome polyprotein [Contains: Matrix protein (Envelope protein M) ;
DE              Major envelope protein E; Nonstructural protein NS1] (Fragment).
OS              Hepatitis C virus (isolate TH) (HCV).
OC              Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC              Hepacivirus.
OC              NCBI_TaxID=11117;
RN              [1]
RP              SEQUENCE FROM N.A.
RX              MEDLINE=91112009; PubMed=1846505;
RA              Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA              Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA              Han J.H.;
RT              "Variable and hypervariable domains are found in the regions of HCV
RT              corresponding to the flavivirus envelope and NS1 proteins and the
RT              pestivirus envelope glycoproteins.";
RL              Virology 180:842-848(1991).
CC              -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC              LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC              PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC              PROTEIN C AND MRNA.

```

D**b** 293 QNIQLINTNG 302

RESULT 3

POLG: HCvH

ID AC PRT; 3011 AA.

AC P27958;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P; Nonstructural protein NS2 (P21) (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

DE Hepatitis C virus (isolate H) (HCV).

OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

OC NCBI_TaxID=11108;

OX [1]

RP SEQUENCE FROM N.A.

RP MEDLINE=92052256; PubMed=1658800;

RX Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Naeoff M., Prince A.M.;

RA "Genomic structure of the human prototype strain H of hepatitis C virus: comparison with American and Japanese isolates.";

RT Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).

RL [2]

RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.

RP MEDLINE=97331322; PubMed=9187654;

RX Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;

RA "Structure of the hepatitis C virus RNA helicase domain.";

RT Nat. Struct. Biol. 4:463-467(1997).

RL [3]

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.

RP MEDLINE=98154321; PubMed=9493270;

RX Kim J.-L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A., Murcko M.A., Lin C., Caron P.R.;

RA "Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding.";

RT Structure 6:89;100(1998).

RL CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.

CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.

CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE ACTIVATION OF NS3.

CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.

CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN ESSENTIAL ROLE IN THE VIRUS REPLICATION.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1 AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.

CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.

CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC

DR EMBL; M67463; AAA45534.1; -.
 DR PIR; A36814; GNVVCH.
 DR PDB; 1HEI; 25-NOV-98.
 DR PDB; 1A1V; 16-FEB-99.
 DR PDB; 1A1R; 17-JUN-98.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR TRANSFAC; T04155; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRP.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR007095; RNA pol Ds ps.
 DR InterPro; IPR007094; RNA pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; helicase C; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 192 383 CAPSID PROTEIN C.
 FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
 FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
 FT CHAIN 810 1026 PROTEIN P7.
 FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
 FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.
 FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4A.
 FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4B.
 FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5A.
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1224 1226

FT TURN 1232 1233
 FT TURN 1236 1238
 FT HELIX 1239 1246
 FT TURN 1247 1248
 FT STRAND 1251 1255
 FT HELIX 1258 1271
 FT TURN 1272 1272
 FT STRAND 1277 1280
 FT TURN 1281 1282
 FT STRAND 1283 1285
 FT STRAND 1291 1295
 FT HELIX 1296 1301
 FT TURN 1302 1303
 FT STRAND 1312 1316
 FT TURN 1317 1319
 FT HELIX 1323 1335
 FT TURN 1336 1340
 FT STRAND 1343 1347
 FT TURN 1352 1353
 FT TURN 1360 1361
 FT STRAND 1362 1366
 FT STRAND 1368 1368
 FT STRAND 1373 1375
 FT TURN 1376 1377
 FT STRAND 1378 1380
 FT TURN 1419 1420
 FT STRAND 1432 1436
 FT TURN 1438 1439
 FT STRAND 1450 1453
 FT STRAND 1456 1463
 FT STRAND 1471 1478
 FT STRAND 1480 1480
 FT HELIX 1481 1488
 FT TURN 1489 1490
 FT STRAND 1497 1501
 FT STRAND 1507 1507
 FT STRAND 1511 1511
 FT HELIX 1514 1527
 FT HELIX 1532 1544
 FT STRAND 1550 1550
 FT HELIX 1555 1564
 FT HELIX 1570 1578
 FT TURN 1579 1580
 FT HELIX 1584 1597
 FT TURN 1598 1598
 FT HELIX 1606 1611
 FT TURN 1614 1618
 FT STRAND 1622 1623
 FT STRAND 1627 1627
 FT STRAND 1635 1636
 FT HELIX 1640 1652
 SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;
 Query Match 50.0%; Score 10; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 0.0066;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QNQLINTNG 10
 DB 409 QNQLINTNG 418
 RESULT 4
 ID POLG HCVH8 STANDARD; PRT; 321 AA.
 AC P27956;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)

16-OCT-2001 (Rel. 40, Last annotation update)
Genome polyprotein [Contains: Matrix protein (Envelope protein M);
Major envelope protein E; Nonstructural protein NS1] (Fragment).
OS Hepatitis C virus (isolate HCT18) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11110;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins.";
RL Virology 180:842-848(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53131; CAA37291.1; -.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT NON TER 1 1
FT CHAIN <1 76
FT CHAIN 268 >321
FT CHAIN 80 80
FT CARBOHYD 93 93
FT CARBOHYD 118 118
FT CARBOHYD 189 189
FT CARBOHYD 301 301
FT CARBOHYD 307 307
FT CARBOHYD 314 314
FT CARBOHYD 321 321
FT NON TER 321 321
SQ SEQUENCE 321 AA; 34238 MW; 2F5DE79F7C7845C8 CRC64;
Query Match 45.0%; Score 9; DB 1; Length 321;
Best Local Similarity 100.0%; Pred. No. 0.01;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 12 WHINSTALN 20
Db 304 WHINSTALN 312
RESULT 5
POLG HCVH7 STANDARD; PRT; 309 AA.
AC P27955;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Matrix protein (Envelope protein M);
DE Major envelope protein E; Nonstructural protein NS1] (Fragment).
OS Hepatitis C virus (isolate HCT27) (HCV).

16-OCT-2001 (Rel. 40, Last annotation update)
Genome polyprotein [Contains: Matrix protein (Envelope protein M);
Major envelope protein E; Nonstructural protein NS1] (Fragment).
OS Hepatitis C virus (isolate HCT18) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11110;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins.";
RL Virology 180:842-848(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53133; CAA37293.1; -.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT NON TER 1 63
FT CHAIN <1 255
FT CHAIN 256 >309
FT CARBOHYD 68 68
FT CARBOHYD 81 81
FT CARBOHYD 106 106
FT CARBOHYD 177 177
FT CARBOHYD 289 289
FT CARBOHYD 295 295
FT CARBOHYD 302 302
FT NON TER 309 309
SQ SEQUENCE 309 AA; 32922 MW; 6E858E9C3D0B9EA9 CRC64;
Query Match 40.0%; Score 8; DB 1; Length 309;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 IQLINTNG 10
Db 283 IQLINTNG 290
RESULT 6
POLG HCVJ2 STANDARD; PRT; 513 AA.
AC P27959;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (Fragment).
OS Hepatitis C virus (isolate HC-J2) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11111;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,

Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111109;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins.";
RL Virology 180:842-848(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53133; CAA37293.1; -.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT NON TER 1 63
FT CHAIN <1 255
FT CHAIN 256 >309
FT CARBOHYD 68 68
FT CARBOHYD 81 81
FT CARBOHYD 106 106
FT CARBOHYD 177 177
FT CARBOHYD 289 289
FT CARBOHYD 295 295
FT CARBOHYD 302 302
FT NON TER 309 309
SQ SEQUENCE 309 AA; 32922 MW; 6E858E9C3D0B9EA9 CRC64;
Query Match 40.0%; Score 8; DB 1; Length 309;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 IQLINTNG 10
Db 283 IQLINTNG 290
RESULT 6
POLG HCVJ2 STANDARD; PRT; 513 AA.
AC P27959;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (Fragment).
OS Hepatitis C virus (isolate HC-J2) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11111;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,

RA Tanaka T., Fukuda S., Tauda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes";
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D10074; BAA00968.1; -;
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 KW INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 >520
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 233 233
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 513 513
 FT NON_TER 513 513
 SQ SEQUENCE 513 AA; 55704 MW; 943F31E3514CDEF3 CRC64;
 Query Match 40.0%; Score 8; DB 1; Length 513;
 Best Local Similarity 100.0%; Pred. No. 0.18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 IQLINTNG 10
 Db 411 IQLINTNG 418
 RESULT 7
 POLG_HCVH4
 ID POLG_HCVH4 STANDARD; PRT; 520 AA.
 AC Q01404;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1)] (Fragment).
 OS Hepatitis C virus (isolate HCV-476) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31643;
 RN [1]
 RP SEQUENCE FROM N.A. PubMed=1383400;
 RX MEDLINE=93019030; PubMed=1383400;
 RA Abe K., Inchauspe G., Fujisawa K.;
 RT "Genomic characterization and mutation rate of hepatitis C virus
 RT isolated from a patient who contracted hepatitis during an epidemic
 RT of non-A, non-B hepatitis in Japan.";
 RL J. Gen. Virol. 73:2725-2729(1992).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D10688; BAA01530.1; -;
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 KW INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 >520
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 418 418
 FT CARBOHYD 424 424
 FT CARBOHYD 431 431
 FT CARBOHYD 449 449
 FT NON_TER 520 520
 SQ SEQUENCE 520 AA; 56499 MW; AAL35246CF20D525 CRC64;
 Query Match 40.0%; Score 8; DB 1; Length 520;
 Best Local Similarity 100.0%; Pred. No. 0.19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 IQLINTNG 10
 Db 412 IQLINTNG 419
 RESULT 8
 POLG_HCVHK
 ID POLG_HCVHK STANDARD; PRT; 520 AA.
 AC Q01403;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1)] (Fragment).

FT	STRAND	1059	1063
FT	STRAND	1068	1074
FT	TURN	1075	1076
FT	TURN	1077	1081
FT	STRAND	1082	1085
FT	TURN	1086	1087
FT	STRAND	1090	1092
FT	TURN	1093	1094
FT	STRAND	1095	1097
FT	TURN	1101	1103
FT	TURN	1104	1107
FT	STRAND	1108	1112
FT	STRAND	1120	1120
FT	STRAND	1122	1122
FT	STRAND	1129	1133
FT	TURN	1135	1136
FT	STRAND	1139	1144
FT	STRAND	1149	1157
FT	TURN	1158	1161
FT	TURN	1162	1163
FT	TURN	1165	1166
FT	STRAND	1168	1171
FT	TURN	1172	1174
FT	STRAND	1175	1186
FT	TURN	1187	1188
FT	STRAND	1189	1197
FT	TURN	1198	1202
FT	TURN	1203	1204
FT	STRAND	1680	1688
SEQ	SEQUENCE	3010 AA; 327189 MW; F8422D5ECCDFD9C CRC64;	

Query Match 40.0%; Score 8; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.93;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	3	IQLINTNG 10
Db	411	IQLINTNG 418

RESULT 10
 POLG_HCVJT
 ID POLG_HCVJT STANDARD; PRT; 3010 AA.
 AC Q00269;
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-JT) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31642;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=92295714; PubMed=1318627;
 RX Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 RT "Molecular cloning of hepatitis C virus genome from a single Japanese
 RT carrier: sequence variation within the same individual and among
 RT infected individuals";
 RL Virus Res. 23:39-53(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys of Thr in P1 and Ser or Ala in P1'.

CC EMBL; M84754; -, NOT_ANNOTATED_CDS.
 DR PIR; A40244; GNVWTW.
 DR PDB; 1N64; 25-FEB-03.
 DR PDB; 1NS3; 08-APR-98.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR INTERPRO; IPR001410; DEAD.
 DR INTERPRO; IPR002522; HCV capsid.
 DR INTERPRO; IPR002521; HCV core.
 DR INTERPRO; IPR002519; HCV env.
 DR INTERPRO; IPR002531; HCV NS1.
 DR INTERPRO; IPR002518; HCV NS2.
 DR INTERPRO; IPR004109; HCV NS3.
 DR INTERPRO; IPR000745; HCV NS4a.
 DR INTERPRO; IPR001490; HCV NS4b.
 DR INTERPRO; IPR002868; HCV NS5a.
 DR INTERPRO; IPR002166; HCV NS5b.
 DR INTERPRO; IPR007095; RNA pol PS.
 DR INTERPRO; IPR007094; RNA pol PSvir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240

FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;
 Query Match 40.0%; Score 8; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.93;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 IQLINTNG 10
 DB 411 IQLINTNG 418
 RESULT 12
 POLG HCVJ6
 ID POLG HCVJ6 STANDARD; PRT; 3033 AA.
 AC P26660;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J6) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 RX NCHI_TaxID=11113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=92044440; PubMed=1658196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 RA Machida A., Miyakawa Y., Mayumi M.;
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 RT from a human carrier: comparison with reported isolates for conserved
 RT and divergent regions.";
 RL J. Gen. Virol. 72:2697-2704(1991).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MENA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D00944; BAA00792.1; -.
 DR PIR; J01303; J01303.
 DR HSSP; P27958; 1HEI.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.


```

DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RGRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA pol Ds Ps.
DR InterPro; IPR007094; RNA pol Psvir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT TRANSMEM 347 369
FT ACT_SITE 1087 1087
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT NP_BIND 1234 1241
FT SITE 1320 1323
FT CARBOHYD 136 196
FT CARBOHYD 209 234
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 578
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT CARBOHYD 1091 1091
FT CARBOHYD 2038 2038
FT CARBOHYD 2811 2811
SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;

Query Match 40.0%; Score 8; DB 1; Length 3033;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 IQLINTNG 10
Db 411 IQLINTNG 418

RESULT 13
PUR9_BUCAP
ID PUR9_BUCAP STANDARD; PRT; 526 AA.

```

```

AC Q8KA70;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bifunctional purine biosynthesis protein purH [Includes:
DE Phosphoribosylamidazolecarboxamide formyltransferase (EC 2.1.2.3)
DE (AICAR transferase); IMP cyclohydrolase (EC 3.5.4.10) (Inosinicase)
DE (IMP synthetase) (ATIC)].
GN PURH OR BUSG032.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tamás I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
RA Wernegreen J.J., Sandstroem J.P., Moran N.A., Andersson S.G.E.;
RA "50 million years of genomic stasis in endosymbiotic bacteria.";
RT Science 296:2376-2379(2002).
CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + 5-amino-1-(5-
CC phospho-D-ribose)imidazole-4-carboxamide = tetrahydrofolate + 5-
CC formamido-1-(5-phospho-D-ribose)imidazole-4-carboxamide.
CC -1- CATALYTIC ACTIVITY: IMP + H(2)O = 5-formamido-1-(5-
CC phosphoribosyl)imidazole-4-carboxamide.
CC -1- PATHWAY: De novo purine biosynthesis; ninth step.
CC -1- PATHWAY: De novo purine biosynthesis; tenth step.
CC -1- DOMAIN: The IMP cyclohydrolase activity resides in the N-terminal
CC region (by similarity).
CC -1- SIMILARITY: BELONGS TO THE PURH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB014079; AAM67603.1; -.
DR HAMAP; MF_00139; -.
DR InterPro; IPR002695; AICARFT_IMPCHas.
DR InterPro; IPR004362; MGS like.
DR Pfam; PF01808; AICARFT_IMPCHas; 1.
DR Pfam; PF02142; MGS; 1.
DR ProDom; PD004666; AICARFT_IMPCHas; 1.
DR TIGRFAMs; TIGR00355; purH; 1.
KW Purine biosynthesis; Transferase; Hydrolase; Multifunctional enzyme;
KW Complete proteome.
SQ SEQUENCE 526 AA; 59421 MW; 7F0F13954EF902A5 CRC64;

Query Match 35.0%; Score 7; DB 1; Length 526;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 IQLINTN 9
Db 157 IQLINTN 163

RESULT 14
POLG_HCVJ5
ID POLG_HCVJ5 STANDARD; PRT; 737 AA.
AC P27960;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
DE protein (Envelope protein M); Major envelope protein E; Nonstructural
DE proteins NS1 and NS2] (Fragment).
DE Hepatitis C virus (isolate HC-J5) (HCV).
OS Hepatitis C virus (isolate HC-J5).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.

```


DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; helicase C; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 3011 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 35.0%; Score 7; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 QLINTNG 10
 |||||
 Db 412 QLINTNG 418

Search completed: November 21, 2003, 20:59:52
 Job time : 5.85 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:45:57 ; Search time 24.4 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-73

Perfect score: 20

Sequence: 1 QNIQLINTNGQWHINSTALN 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	10	50.0	58	12	Q96875 hepatitis c
2	10	50.0	58	12	Q96874 hepatitis c
3	10	50.0	64	12	Q91JW4 hepatitis c
4	10	50.0	64	12	Q91JW9 hepatitis c
5	10	50.0	66	12	Q68559 hepatitis c
6	10	50.0	69	12	Q72305 hepatitis c
7	10	50.0	74	12	Q9WLK7 hepatitis c
8	10	50.0	75	12	Q9W916 hepatitis c
9	10	50.0	76	12	Q92982 hepatitis c
10	10	50.0	82	12	Q72244 hepatitis c
11	10	50.0	85	12	Q91589 hepatitis c
12	10	50.0	85	12	Q915U0 hepatitis c
13	10	50.0	85	12	Q915R2 hepatitis c
14	10	50.0	85	12	Q915U2 hepatitis c
15	10	50.0	85	12	Q915U3 hepatitis c
16	10	50.0	85	12	Q91586 hepatitis c

17	10	50.0	85	12	Q915T0 hepatitis c
18	10	50.0	85	12	Q915T6 hepatitis c
19	10	50.0	85	12	Q915T8 hepatitis c
20	10	50.0	85	12	Q915T9 hepatitis c
21	10	50.0	85	12	Q915T4 hepatitis c
22	10	50.0	85	12	Q915T3 hepatitis c
23	10	50.0	85	12	Q915T1 hepatitis c
24	10	50.0	85	12	Q915T7 hepatitis c
25	10	50.0	85	12	Q915U1 hepatitis c
26	10	50.0	85	12	Q915S8 hepatitis c
27	10	50.0	85	12	Q915T5 hepatitis c
28	10	50.0	85	12	Q915U4 hepatitis c
29	10	50.0	85	12	Q915S7 hepatitis c
30	10	50.0	90	12	P89661 hepatitis c
31	10	50.0	91	12	Q9WG67 hepatitis c
32	10	50.0	93	12	Q72259 hepatitis c
33	10	50.0	96	12	Q91ZNS hepatitis c
34	10	50.0	97	12	Q8JK94 hepatitis c
35	10	50.0	99	12	Q8JKA0 hepatitis c
36	10	50.0	100	12	Q8QRI3 hepatitis c
37	10	50.0	104	12	Q72260 hepatitis c
38	10	50.0	105	12	Q72261 hepatitis c
39	10	50.0	106	12	Q72256 hepatitis c
40	10	50.0	106	12	Q72257 hepatitis c
41	10	50.0	106	12	Q72254 hepatitis c
42	10	50.0	106	12	Q72255 hepatitis c
43	10	50.0	109	12	Q91JW7 hepatitis c
44	10	50.0	109	12	Q91JW3 hepatitis c
45	10	50.0	113	12	Q9E725 hepatitis c

ALIGNMENTS

RESULT 1

Q96875 ID Q96875 PRELIMINARY; PRT; 58 AA.
AC Q96875;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Envelope glycoprotein E2/NS1 (Genome polyprotein) (Fragment).
GN E2/NS1.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand-viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_taxid=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HO7;
RX MEDLINE=96416052; PubMed=8818959;
RA Yun Z., Lara C., Johansson B., Lorenzana de Rivera I., Sonnerborg A.;
RT "Discrepancy of hepatitis C virus genotypes as determined by
phylogenetic analysis of partial NS5 and core sequences.";
J. Med. Virol. 49:155-160(1996).
DR EMBL; U28037; AAC55458.1; -;
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 58
SQ SEQUENCE 58 AA; 6068 MW; E426BFB6BD24B3CC CRC64;

Query Match 50.0%; Score 10; DB 12; Length 58;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
QNIQLINTNG 10
|||||||

```
Db 41 QNIQLINTNG 50
RESULT 2
Q96874 ID Q96874 PRELIMINARY; PRT; 58 AA.
AC Q96874;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Envelope glycoprotein E2/NS1 (Genome polyprotein) (Fragment).
GN E2/NS1.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H06;
RX MEDLINE=96416052; PubMed=8818959;
RA Yun Z., Lara C., Johanson B., Lorenzana de Rivera I., Sonnerborg A.;
RT "Discrepancy of hepatitis C virus genotypes as determined by
RT phylogenetic analysis of partial NS5 and core sequences.";
RL J. Med. Virol. 49:155-160(1996).
DR EMBL; U28036; AAC55457.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 58
SQ SEQUENCE 58 AA; 6082 MW; C48DBE361724D244 CRC64;

Query Match 50.0%; Score 10; DB 12; Length 58;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
Db 41 QNIQLINTNG 50

RESULT 3
Q91JW4 ID Q91JW4 PRELIMINARY; PRT; 64 AA.
AC Q91JW4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Isolate hc41 E2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=hc41;
RX "Genetic Characteristics of Putative Hypervariable Region (HVR1) in
RT Genotype 2 Hepatitis C virus.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF218875; AAF80572.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 58
SQ SEQUENCE 58 AA; 6082 MW; C48DBE361724D244 CRC64;

Query Match 50.0%; Score 10; DB 12; Length 58;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
Db 41 QNIQLINTNG 50

RESULT 4
Q91JW9 ID Q91JW9 PRELIMINARY; PRT; 64 AA.
AC Q91JW9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Isolate hc21 E2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=hc21;
RX Fan X., Di Bisceglie A.M.;
RT "Genetic Characteristics of Putative Hypervariable Region (HVR1) in
RT Genotype 2 Hepatitis C virus.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF218870; AAF80567.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 64
SQ SEQUENCE 64 AA; 6949 MW; 8BD6E30BD182C88C CRC64;

Query Match 50.0%; Score 10; DB 12; Length 64;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
Db 51 QNIQLINTNG 60

RESULT 5
Q68559 ID Q68559 PRELIMINARY; PRT; 66 AA.
AC Q68559;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Envelope glycoprotein E2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GamGrd pattern A;
RX MEDLINE=97240215; PubMed=9122906;
RA Widell A., Zhang Y.Y., Andersson-Gare B., Hammarstrom L.;
RT "At least three hepatitis C virus strains implicated in Swedish and
RT Danish patients with intravenous immunoglobulin-associated hepatitis
RT C.";
RL Transfusion 37:313-320(1997).
```

DR EMBL; U51791; AAB71435.1; -;
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1 1
 FT NON_TER 66 66
 SQ SEQUENCE 66 AA; 6895 MW; D320BE43CDB04FID CRC64;
 Query Match 50.0%; Score 10; DB 12; Length 66;
 Best Local Similarity 100.0%; Pred. No. 0.0014;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QNIQLINTNG 10
 Db |||||
 32 QNIQLINTNG 41
 RESULT 6
 O72305 PRELIMINARY; PRT; 69 AA.
 AC O72305;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RC STRAIN=anti-D Recipient 68;
 EX MEDLINE=98241727; PubMed=9573256;
 RA McAllister J., Casino C., Davidson F., Power J., Lawlor E., Yap P.L.,
 RA Simmonds P., Smith D.B.;
 RT "Long-term evolution of the hypervariable region of hepatitis C virus
 in a common-source-infected cohort.";
 RL J. Virol. 72:4893-4905(1998).
 DR EMBL; AF056883; AAC17379.1; -;
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1 1
 FT NON_TER 69 69
 SQ SEQUENCE 69 AA; 7357 MW; F257BCCF7E4BFFAE CRC64;
 Query Match 50.0%; Score 10; DB 12; Length 69;
 Best Local Similarity 100.0%; Pred. No. 0.0014;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QNIQLINTNG 10
 Db |||||
 31 QNIQLINTNG 40
 RESULT 7
 O9WLK7 PRELIMINARY; PRT; 74 AA.
 AC O9WLK7;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RC STRAIN=anti-D Recipient 68;
 EX MEDLINE=98241727; PubMed=9573256;
 RA McAllister J., Casino C., Davidson F., Power J., Lawlor E., Yap P.L.,
 RA Simmonds P., Smith D.B.;
 RT "Long-term evolution of the hypervariable region of hepatitis C virus
 in a common-source-infected cohort.";
 RL J. Virol. 72:4893-4905(1998).
 DR EMBL; AF056883; AAC17379.1; -;
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1 1
 FT NON_TER 69 69
 SQ SEQUENCE 69 AA; 7357 MW; F257BCCF7E4BFFAE CRC64;
 Query Match 50.0%; Score 10; DB 12; Length 69;
 Best Local Similarity 100.0%; Pred. No. 0.0014;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QNIQLINTNG 10
 Db |||||
 31 QNIQLINTNG 40
 RESULT 7
 O9WLK7 PRELIMINARY; PRT; 74 AA.
 AC O9WLK7;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RC STRAIN=anti-D Recipient 68;
 EX MEDLINE=98241727; PubMed=9573256;
 RA McAllister J., Casino C., Davidson F., Power J., Lawlor E., Yap P.L.,
 RA Simmonds P., Smith D.B.;
 RT "Long-term evolution of the hypervariable region of hepatitis C virus
 in a common-source-infected cohort.";
 RL J. Virol. 72:4893-4905(1998).
 DR EMBL; AF056883; AAC17379.1; -;
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1 1
 FT NON_TER 69 69
 SQ SEQUENCE 69 AA; 7357 MW; F257BCCF7E4BFFAE CRC64;
 Query Match 50.0%; Score 10; DB 12; Length 69;
 Best Local Similarity 100.0%; Pred. No. 0.0014;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QNIQLINTNG 10
 Db |||||
 31 QNIQLINTNG 40

RC STRAIN=anti-D Recipient 68;
 RX MEDLINE=98241727; PubMed=9573256;
 RA McAllister J., Casino C., Davidson F., Power J., Lawlor E., Yap P.L.,
 RA Simmonds P., Smith D.B.;
 RT "Long-term evolution of the hypervariable region of hepatitis C virus
 in a common-source-infected cohort.";
 RL J. Virol. 72:4893-4905(1998).
 DR EMBL; AF056885; AAC17381.1; -;
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1 1
 FT NON_TER 74 74
 SQ SEQUENCE 74 AA; 7777 MW; 89933164D292B86F CRC64;
 Query Match 50.0%; Score 10; DB 12; Length 74;
 Best Local Similarity 100.0%; Pred. No. 0.0015;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QNIQLINTNG 10
 Db |||||
 30 QNIQLINTNG 39
 RESULT 8
 O9W9I6 PRELIMINARY; PRT; 75 AA.
 AC O9W9I6;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RC STRAIN=anti-D Recipient 68;
 EX MEDLINE=98241727; PubMed=9573256;
 RA McAllister J., Casino C., Davidson F., Power J., Lawlor E., Yap P.L.,
 RA Simmonds P., Smith D.B.;
 RT "Long-term evolution of the hypervariable region of hepatitis C virus
 in a common-source-infected cohort.";
 RL J. Virol. 72:4893-4905(1998).
 DR EMBL; AF056889; AAC17385.1; -;
 DR EMBL; AF056884; AAC17380.1; -;
 DR EMBL; AF056886; AAC17382.1; -;
 DR EMBL; AF056888; AAC17384.1; -;
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1 1
 FT NON_TER 75 75
 SQ SEQUENCE 75 AA; 7848 MW; 89936131D7C2ED6A CRC64;
 Query Match 50.0%; Score 10; DB 12; Length 75;
 Best Local Similarity 100.0%; Pred. No. 0.0015;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QNIQLINTNG 10
 Db |||||
 31 QNIQLINTNG 40
 RESULT 9
 O92982 PRELIMINARY; PRT; 76 AA.
 ID O92982
 AC O92982;

DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein. (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=anti-D Recipient 68;
 RX MEDLINE=98241727; PubMed=9573256;
 RA McAllister J., Casino C., Davidson F., Power J., Lawlor E., Yap P.L.,
 RA Simmonds P., Smith D.B.;
 RT "Long-term evolution of the hypervariable region of hepatitis C virus
 in a common-source-infected cohort.";
 RL J. Virol. 72:4893-4905(1998).
 DR EMBL; AF056887; AAC17383.1; -.
 DR InterPro; IPR002531; HCV_NSI.
 DR Pfam; PF01560; HCV_NSI; 1.
 DR ProDom; PD186062; HCV_NSI; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 76
 FT NON_TER 76
 SQ SEQUENCE 76 AA; 7995 MW; 2EB361675F1CBBE2 CRC64;

 Query Match 50.0%; Score 10; DB 12; Length 76;
 Best Local Similarity 100.0%; Pred. No. 0.0015;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 Qy 1 QNIQLINTNG 10
 Db |||||
 Db 32 QNIQLINTNG 41

 RESULT 10
 ID 072244 PRELIMINARY; PRT; 82 AA.
 AC 072244;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein. (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=anti-D Recipient 12;
 RX MEDLINE=98241727; PubMed=9573256;
 RA McAllister J., Casino C., Davidson F., Power J., Lawlor E., Yap P.L.,
 RA Simmonds P., Smith D.B.;
 RT "Long-term evolution of the hypervariable region of hepatitis C virus
 in a common-source-infected cohort.";
 RL J. Virol. 72:4893-4905(1998).
 DR EMBL; AF056812; AAC17308.1; -.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV_NSI.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV_NSI; 1.
 DR ProDom; PD186062; HCV_NSI; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 82
 FT NON_TER 82
 SQ SEQUENCE 82 AA; 8814 MW; 0AACFECC7303DEDE CRC64;

 Query Match 50.0%; Score 10; DB 12; Length 82;
 Best Local Similarity 100.0%; Pred. No. 0.0016;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10
 Db |||||
 Db 44 QNIQLINTNG 53

 RESULT 11
 ID 0915S9 PRELIMINARY; PRT; 85 AA.
 AC 0915S9;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Glycoprotein (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Ball J.K., Curran R., Jameson C.L., Craggs J.K., Grabowska A.M.,
 RA Thomson B.J., Robins A., Irving W.L., Ball J.K.;
 RT "Evolutionary trends of the first hypervariable region of the
 hepatitis C virus E2 protein in individuals with differing liver
 disease severity.";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF426589; AAL24935.1; -.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV_NSI.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV_NSI; 1.
 DR ProDom; PD186062; HCV_NSI; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 85
 FT NON_TER 85
 SQ SEQUENCE 85 AA; 8862 MW; C35605D3F064188B CRC64;

 Query Match 50.0%; Score 10; DB 12; Length 85;
 Best Local Similarity 100.0%; Pred. No. 0.0017;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 Qy 1 QNIQLINTNG 10
 Db |||||
 Db 74 QNIQLINTNG 83

 RESULT 12
 ID 0915U0 PRELIMINARY; PRT; 85 AA.
 AC 0915U0;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Glycoprotein (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Ball J.K., Curran R., Jameson C.L., Craggs J.K., Grabowska A.M.,
 RA Thomson B.J., Robins A., Irving W.L., Ball J.K.;
 RT "Evolutionary trends of the first hypervariable region of the
 hepatitis C virus E2 protein in individuals with differing liver
 disease severity.";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF426578; AAL24924.1; -.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV_NSI.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV_NSI; 1.
 DR ProDom; PD186062; HCV_NSI; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;

KW Polypeptide; Transmembrane.

FT NON_TER 1

FT NON_TER 85

SQ SEQUENCE 85 AA; 8840 MW; DBDB5D043BE7BA7B CRC64;

Query Match 50.0%; Score 10; DB 12; Length 85;

Best Local Similarity 100.0%; Pred. No. 0.0017;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10

|||||

Db 74 QNIQLINTNG 83

RESULT 13

ID Q915T2

AC Q915T2 PRELIMINARY; PRT; 85 AA.

DT 01-DEC-2001 (TReMBLrel. 19, Created)

DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)

DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)

DE Glycoprotein (Genome polypeptide) (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11103;

RN [1]

RP SEQUENCE FROM N.A.

RA Ball J.K., Curran R., Jameson C.L., Craggs J.K., Grabowska A.M.,

RA Thomson B.J., Robins A., Irving W.L., Ball J.K.;

RT "Evolutionary trends of the first hypervariable region of the

RT hepatitis C virus E2 protein in individuals with differing liver

RT disease severity."

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF426586; AAL24932.1; -

DR InterPro; IPR002519; HCV env.

DR InterPro; IPR002531; HCV NS1.

DR Pfam; PF01539; HCV env; 1.

DR Pfam; PF01560; HCV NS1; 1.

DR ProDom; PD186062; HCV NS1; 1.

KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;

KW Polypeptide; Transmembrane.

FT NON_TER 1

FT NON_TER 85

SQ SEQUENCE 85 AA; 8820 MW; C356166BD0BEA93B CRC64;

Query Match 50.0%; Score 10; DB 12; Length 85;

Best Local Similarity 100.0%; Pred. No. 0.0017;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10

|||||

Db 74 QNIQLINTNG 83

RESULT 14

ID Q915U2

AC Q915U2 PRELIMINARY; PRT; 85 AA.

DT 01-DEC-2001 (TReMBLrel. 19, Created)

DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)

DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)

DE Glycoprotein (Genome polypeptide) (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11103;

RN [1]

RP SEQUENCE FROM N.A.

RA Ball J.K., Curran R., Jameson C.L., Craggs J.K., Grabowska A.M.,

RA Thomson B.J., Robins A., Irving W.L., Ball J.K.;

RT "Evolutionary trends of the first hypervariable region of the

RT hepatitis C virus E2 protein in individuals with differing liver

RT disease severity."

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF426576; AAL24922.1; -

DR InterPro; IPR002519; HCV env.

DR InterPro; IPR002531; HCV NS1.

DR Pfam; PF01539; HCV env; 1.

DR Pfam; PF01560; HCV NS1; 1.

KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;

KW Polypeptide; Transmembrane.

FT NON_TER 1

FT NON_TER 85

SQ SEQUENCE 85 AA; 8832 MW; DBC78CC923A0E57B CRC64;

Query Match 50.0%; Score 10; DB 12; Length 85;

Best Local Similarity 100.0%; Pred. No. 0.0017;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10

|||||

Db 74 QNIQLINTNG 83

RESULT 15

ID Q915U3

AC Q915U3 PRELIMINARY; PRT; 85 AA.

DT 01-DEC-2001 (TReMBLrel. 19, Created)

DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)

DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)

DE Glycoprotein (Genome polypeptide) (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11103;

RN [1]

RP SEQUENCE FROM N.A.

RA Ball J.K., Curran R., Jameson C.L., Craggs J.K., Grabowska A.M.,

RA Thomson B.J., Robins A., Irving W.L., Ball J.K.;

RT "Evolutionary trends of the first hypervariable region of the

RT hepatitis C virus E2 protein in individuals with differing liver

RT disease severity."

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF426575; AAL24921.1; -

DR InterPro; IPR002519; HCV env.

DR InterPro; IPR002531; HCV NS1.

DR Pfam; PF01539; HCV env; 1.

DR Pfam; PF01560; HCV NS1; 1.

KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;

KW Polypeptide; Transmembrane.

FT NON_TER 1

FT NON_TER 85

SQ SEQUENCE 85 AA; 8832 MW; DBC78CC923A0E57B CRC64;

Query Match 50.0%; Score 10; DB 12; Length 85;

Best Local Similarity 100.0%; Pred. No. 0.0017;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNIQLINTNG 10

|||||

Db 74 QNIQLINTNG 83

Search completed: November 21, 2003, 21:08:12

Job time : 25.45 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:22 ; Search time 33.705 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-74

Perfect score: 21

Sequence: 1 LNCNLSNTGWLGLYQHK 21

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq 19Jun03.*

- 1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	21	16	AA1984503
2	21	100.0	21	17	AA1990997
3	21	100.0	21	23	AAO18703
4	11	52.4	27	24	ABP70586
5	11	52.4	30	23	AAU84626
6	11	52.4	40	13	AA1920760
7	11	52.4	91	23	ABP77256
8	11	52.4	144	14	AA193998
9	11	52.4	174	14	AA193441

10	11	52.4	250	23	AAU79220	Hepatitis C Virus
11	11	52.4	271	22	AAE00445	HCV E2-634 HVR1 CO
12	11	52.4	271	22	AAE00446	HCV E2-634 HVR1-mu
13	11	52.4	305	18	AAW00929	Recombinant HCV E2
14	11	52.4	319	15	AAW45330	Anti-HCV antibody
15	11	52.4	333	14	AA1940118	HGH-HCV-E2 fusion
16	11	52.4	337	16	AA1979217	PHCV351-encoded AP
17	11	52.4	350	23	AAE19891	Hepatitis C virus
18	11	52.4	363	24	ABP55670	Hepatitis C virus
19	11	52.4	363	24	ABP55671	Hepatitis C virus
20	11	52.4	367	14	AA1940115	APP-HCV-E2 fusion
21	11	52.4	367	16	AA1979218	PHCV167-encoded pr
22	11	52.4	377	16	AA1979226	PHCV422-encoded AP
23	11	52.4	397	16	AA1979220	PHCV419-encoded AP
24	11	52.4	399	14	AA1940117	HGH-HCV-E2 fusion
25	11	52.4	402	14	AA1934439	Sequence of glycop
26	11	52.4	409	14	AA1933995	H77 E2/NS1 protein
27	11	52.4	410	16	AA1979227	PHCV423-encoded AP
28	11	52.4	417	16	AA1979228	PHCV424-encoded AP
29	11	52.4	434	16	AA1979219	PHCV418-encoded AP
30	11	52.4	441	16	AA1979230	PHCV429-encoded AP
31	11	52.4	447	16	AA1979229	PHCV425-encoded AP
32	11	52.4	453	16	AA1979225	PHCV421-encoded AP
33	11	52.4	467	13	AA1921575	HCV CKS-3'ENV - pH
34	11	52.4	467	14	AA1933642	HCV CKS-3'ENV reco
35	11	52.4	467	14	AA1933584	HCV CKS-3'ENV reco
36	11	52.4	467	14	AA1933604	HCV CKS-3'ENV reco
37	11	52.4	467	22	AA1969003	HCV recombinant an
38	11	52.4	490	16	AA1979224	PHCV420-encoded AP
39	11	52.4	497	22	AAE00443	HCV HVR1-7 constru
40	11	52.4	497	22	AAE00444	HCV HVR1-mut5 cons
41	11	52.4	502	16	AA197589	Hepatitis C virus
42	11	52.4	513	13	AA1924086	NANB hepatitis vir
43	11	52.4	527	13	AA1925136	HCV polypeptide 2
44	11	52.4	537	14	AA1940114	APP-HCV-E2 fusion
45	11	52.4	733	14	AA1938278	NANB hepatitis vir

ALIGNMENTS

RESULT 1
AA1984503
ID AA1984503 standard; peptide; 21 AA.
AC AA1984503;
XX
XX
DT 06-JAN-1997 (first entry)
DE Hepatitis C virus peptide NS1-3* (residues 427-446).
DE Hepatitis C virus; HCV; immunogen; non-structural region; NS1;
KW immunodominant; T cell epitope; vaccine.
XX
XX
OS Hepatitis C virus.
PN WO9512677-A2.
XX
PD 11-MAY-1995.
XX
PF 28-OCT-1994; 94WO-EP03555.
XX
PR 04-NOV-1993; 93EP-0402718.
XX
XX (INNO-) INNOGENETICS NV.
PA Deleys R, Leroux-Roels G, Maertens G;
PI WPI; 1995-193822/25.
DR Hepatitis C Virus immunogenic polypeptide contg. a T-cell
PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in
PT production of vaccines, therapeutic agents, etc.

XX Example 4; Page 51; 105pp; English.

XX A series of overlapping peptides (including the present sequence) was

CC synthesised based on sequences in the core, E1 and E2/NS1 regions of

CC hepatitis C virus. The peptides were used as antigens in lympho-

CC proliferative assays to identify the main T-cell epitopes.

XX Sequence 21 AA;

SQ Query Match 100.0%; Score 21; DB 16; Length 21;

Best Local Similarity 100.0%; Pred. No. 4.1e-15;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGWLGLAGLIYQHK 21

Db 1 LNCNESLNTGWLGLAGLIYQHK 21

RESULT 2

AAR90997

ID AAR90997 standard; peptide; 21 AA.

XX AC AAR90997;

XX DT 25-SEP-1996 (first entry)

XX DE HCV E2 peptide E2-3B for competition studies.

XX KW HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;

XX KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.

XX OS Synthetic.

XX PN WO9604385-A2.

XX PD 15-FEB-1996.

XX PF 31-JUL-1995; 95WO-EP03031.

XX PR 29-JUL-1994; 94EP-0870132.

XX PA (INNO-) INNOGENETICS NV.

XX PI Bosman F, Buyse M, De Martynoff G, Maertens G;

XX WPI; 1996-129401/13.

XX Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope

PT proteins - in presence of disulphide bond cleavage agent, to

PT produce proteins suitable for direct use in vaccines or diagnostic

PT assays of HCV

XX Example 7; Page 67; 146pp; English.

XX AAR90965-R90980 and AAR90995-R91015 represent synthetic hepatitis C

CC virus (HCV) E1 and E2 peptides used in competition studies. This

CC sequence represents a synthetic E2 peptide, and corresponds to residues

CC 427-446 of the E2 protein sequence. These sequences are useful for in

CC vitro monitoring of HCV disease, or prognosis of the response to

CC interferon treatment of patients suffering from HCV infection. These

CC sequences compete with the proteins produced by AAT12704-T12709 and

CC AAT12961-T12974, which are included in vectors for the production of

CC recombinant E1, E2, and E1/E2. The recombinant proteins can then be

CC isolated and purified by carrying out a disulphide bond cleavage, or a

CC reduction step with a disulphide bond cleavage agent, after lysis of

CC recombinant host cells. The constructs containing the purified HCV

CC envelope proteins can be used for vaccinating humans against HCV, for in

CC vitro detection of HCV antibodies in a sample, and in a serotyping assay

CC for detecting one or more serological types of HCV present in a

CC biological sample. The constructs can also be immobilised on a solid

CC substrate and incorporated into a reversed phase hybridisation assay for

CC determining the presence or the genotype of HCV. The new purification

CC method preserves the conformation of the recombinantly expressed E1, E2

CC and E1/E2, and eliminates contaminating proteins. Antigens isolated

CC using this method are more reactive with human sera than those isolated

CC by known techniques.

XX Sequence 21 AA;

SQ Query Match 100.0%; Score 21; DB 17; Length 21;

Best Local Similarity 100.0%; Pred. No. 4.1e-15;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGWLGLAGLIYQHK 21

Db 1 LNCNESLNTGWLGLAGLIYQHK 21

RESULT 3

AAO18703

ID AAO18703 standard; Peptide; 21 AA.

XX AC AAO18703;

XX DT 24-OCT-2002 (first entry)

XX DE Hepatitis C virus E2 protein derived peptide E2-3B.

XX KW Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide;

XX KW immunostimulant; vaccine.

XX OS Hepatitis C virus.

XX PN WO200255548-A2.

XX PD 18-JUL-2002.

XX PF 11-JAN-2002; 2002WO-EP00219.

XX PR 11-JAN-2001; 2001US-260699P.

XX PR 30-AUG-2001; 2001US-315768P.

XX PA (INNO-) INNOGENETICS NV.

XX PI Maertens G, Bosman F, Buyse M;

XX WPI; 2002-599657/64.

XX New therapeutic vaccine compositions comprising at least one purified

PT recombinant hepatitis C virus (HCV) single or specific oligomeric

PT recombinant envelope protein E1 or E2, useful for immunizing humans

PT from HCV infection

XX Example 7; Page 226; 243pp; English.

XX The present invention relates to new therapeutic vaccine compositions for

CC inducing hepatitis C virus (HCV)-specific antibodies, comprising a

CC composition containing at least one purified recombinant HCV single or

CC specific oligomeric recombinant envelope proteins selected from an E1 and

CC an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are

CC useful for inducing HCV-specific antibodies or for immunising humans

CC against HCV. The recombinant HCV E1 and/or E2 proteins are useful as

CC vaccines or therapeutics, in HCV screening and confirmatory antibody

CC tests for raising antibodies, in the preparation of medicament, and for

CC in vitro monitoring of HCV disease or prognosing the response to

CC treatment of patients suffering from HCV infection. The present sequence

CC is a peptide derived from the proteins of the invention.

XX Sequence 21 AA;

SQ Query Match 100.0%; Score 21; DB 23; Length 21;

Best Local Similarity 100.0%; Pred. No. 4.1e-15;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGWLGLAGLIYQHK 21

Db 1 LNCNESLNTGWWLAGLIYQHK 21
|||||

RESULT 4

ABP70586
ID ABP70586 standard; peptide; 27 AA.

XX AC ABP70586;
XX DT 22-APR-2003 (first entry)

XX DE Hepatitis C virus E2 glycoprotein derived antigenic peptide.
XX KW Flavivirus; vector; envelope protein; peptide delivery; immune response;
XX KW tumour-associated antigen; immunotherapy; cancer; HCV; E2 glycoprotein.

XX OS Hepatitis C virus.
XX XX WO2002102828-A2.
XX PD 27-DEC-2002.

XX PF 31-MAY-2002; 2002WO-US17374.
XX PR 01-JUN-2001; 2001US-295265P.
XX PA (ACAM-) ACAMBIS INC.

XX PI Kleanthous H, Oros L, Miller C;
XX DR WPI; 2003-167480/16.
XX PT Novel flavivirus vector useful for delivering foreign peptides,
XX PS comprises an envelope protein that comprises a foreign peptide
XX Example; Page 18; 34pp; English.

XX CC The specification describes a flavivirus vector, comprising an envelope
XX CC protein that comprises a foreign peptide. The vector is useful for
XX CC delivery of a peptide, e.g. an antigen to a patient. It can be
XX CC administered to induce an immune response to a pathogen or tumour from
XX CC which the antigen is derived. It also useful for delivering
XX CC tumour-associated antigens for use in immunotherapeutic methods against
XX CC cancer. ABP70571-96 represent B and T cell epitopes from viruses.
XX CC They may be expressed using the vectors of the invention.
XX SQ Sequence 27 AA;

Query Match 52.4%; Score 11; DB 24; Length 27;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
Db 17 LNCNESLNTGW 27
|||||

RESULT 5

AAU84626
ID AAU84626 standard; Peptide; 30 AA.

XX AC AAU84626;
XX DT 08-MAY-2002 (first entry)

XX DE HCV HepC1a segment 29.
XX KW Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
XX KW viral infection; human immunodeficiency virus; melanoma;
XX KW bacterial infection; Salmonella; Legionella; parasitic infection;
XX KW Trypanosoma; Toxoplasma; Giardia.

OS Hepatitis C virus.
XX PN WO200190197-A1.
XX PD 29-NOV-2001.
XX PF 25-MAY-2001; 2001WO-AU00622.
XX PR 26-MAY-2000; 2000AU-0007761.
XX PA (AUSU) UNIV AUSTRALIAN NAT.
XX PI Thomson SA, Ramshaw IA;
XX DR WPI; 2002-147575/19.
XX DR N-PSDB; ABK36464.

XX FT New synthetic polypeptides having several different segments of at
XX PT least one parent polypeptide linked together differently compared to
XX PT the linkage in the parent polypeptide, for inducing immune response
XX PT against a pathogen or cancer
XX PS Example 2; Fig 26; 364pp; English.

XX CC The invention relates to a new synthetic polypeptide (I) comprising
XX CC several different segments of at least one parent polypeptide linked
XX CC together in a different relationship relative to their linkage in the
XX CC parent polypeptide to impede, abrogate or otherwise alter at least one
XX CC function associated with the parent polypeptide and for inducing an
XX CC immune response against a pathogen or cancer. Also included are a
XX CC synthetic polynucleotide encoding and a computer system for
XX CC designing the synthetic polypeptides. The synthetic polypeptide is
XX CC polynucleotides are referred to as a Savine. The synthetic polypeptide is
XX CC useful for modulating immune responses preferably directed against a
XX CC pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix,
XX CC colon, head and neck, pancreas, prostate, stomach, kidney, bone
XX CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.
XX CC Compositions comprising the polypeptide may be used in the treatment or
XX CC prophylaxis against viral (such as infections caused by HIV (human
XX CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
XX CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
XX CC (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,
XX CC Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic
XX CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
XX CC Trypanosoma, Toxoplasma and Giardia) infections. The present
XX CC sequence is a peptide derived from a parent protein used to
XX CC construct a savine of the invention.
XX SQ Sequence 30 AA;

Query Match 52.4%; Score 11; DB 23; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
Db 9 LNCNESLNTGW 19
|||||

RESULT 6
AAR20760
ID AAR20760 standard; Protein; 40 AA.

XX AC AAR20760;
XX DT 25-MAR-2003 (updated)
XX DT 05-MAY-1992 (first entry)

XX DE Peptide 10 based on immunoreactive region of Hepatitis C virus.
XX KW Non-A, non-B hepatitis virus; non-structural protein; vaccine.
XX XX Synthetic.

XX EP468527-A.
 XX 29-JAN-1992.
 XX 26-JUL-1991; 91EP-0112620.
 XX 24-JUN-1991; 91US-0719819.
 XX 26-JUL-1990; 90US-0558799.
 XX 07-FEB-1991; 91US-0651735.
 XX 11-MAR-1991; 91US-0667275.
 XX (UNBI-) UNITED BIOMEDICAL INC.
 XX Chang YW, Hosein B;
 XX WPI; 1992-034279/05.
 XX New synthetic peptide specific for HCV antibodies - for detection
 PT of HCV or NANBHV e.g. by enzyme-linked immunosorbent assay and is
 PT immunogen for preparation of vaccines
 XX Claim 1; Page 89; 98pp; English.
 XX This peptide is one of 19 specifically claimed antigens based on the
 CC immunoreactive regions of the envelope protein and non-structural
 CC proteins NS-1, NS-2, NS-3 and NS-5 for the Hepatitis C virus. The
 CC invention also concerns analogues, segments, mixtures, conjugates
 CC and polymers of these peptides. The C-terminal amino acid may be
 CC amidated. See AAR20751-R20782.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 XX Sequence 40 AA;
 SQ

Query Match 52.4%; Score 11; DB 13; Length 40;
 Best Local Similarity 100.0%; Pred. No. 0.00017;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
 |||||
 Db 8 LNCNESLNTGW 18

RESULT 7
 ABB77256
 ID ABB77256 standard; Protein; 91 AA.
 XX AC ABB77256;
 XX 28-JUN-2002 (first entry)
 XX HCV bait polypeptide 4.
 XX SID; selected interacting domain; HCV; hepatitis C virus;
 XX liver disease; liver cancer; virucide; hepatotropic; antiinflammatory;
 XX antibacterial.
 XX Hepatitis C virus strain H77.
 XX EP1178116-A1.
 XX 06-FEB-2002.
 XX 03-AUG-2000; 2000EP-0402225.
 XX 03-AUG-2000; 2000EP-0402225.
 XX (HYBR-) HYBRIGENICS SA.
 XX Legrain P, Whiteside S, Wojcik J;
 XX WPI; 2002-208115/27.
 XX N-PSDB; ABL55588.
 DR

XX New selected interacting domain polypeptides and polynucleotides,
 PT useful for treating or preventing infections or pathologies caused by
 PT hepatitis C virus (HCV) or those linked to HCV infection -
 XX Claim 26; SEQ ID 80; 61pp + sequence listing; English.
 XX The invention relates to nucleic acids encoding polypeptides which are
 CC termed SID polypeptides (selected interacting domain). These polypeptides
 CC are the final products of a double selection method involving a first
 CC step of selection of Hepatitis C virus (HCV)-derived polynucleotides
 CC through a two-hybrid system, and a second selection step involving an
 CC alignment between the different polynucleotides selected at the first
 CC step. The activity of polypeptides of the invention may be described as,
 CC virucide, hepatotropic, antiinflammatory and antibacterial. The
 CC polypeptide, polynucleotide and compositions comprising them are useful
 CC for treating or preventing viral or a bacterial infection, specifically
 CC infections or pathologies caused by HCV, or those pathologies linked to
 CC HCV infection. These may include liver disease and liver cancer. The
 CC current sequence represents a HCV bait polypeptide.
 CC Note: The sequence data for this patent is not represented in the
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.
 XX SQ Sequence 91 AA;
 Query Match 52.4%; Score 11; DB 23; Length 91;
 Best Local Similarity 100.0%; Pred. No. 0.00035;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
 |||||
 Db 48 LNCNESLNTGW 58

RESULT 8
 AAR33998
 ID AAR33998 standard; Protein; 144 AA.
 XX AC AAR33998;
 XX 25-MAR-2003 (updated)
 XX 26-JUL-1993 (first entry)
 XX HC-J1 E2/NS1 protein.
 XX Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
 XX HCV; asymptomatic; chronically infected; epitope; viral isolate;
 XX domain; immunological; cross-reactive; envelope protein; vaccine;
 XX gp53 (BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.
 XX Synthetic.
 XX WO9306126-A1.
 XX 01-APR-1993.
 XX 11-SEP-1992; 92WO-US07683.
 XX 13-SEP-1991; 91US-0759575.
 XX (CHIR) CHIRON CORP.
 XX Houghton M, Weiner AJ;
 XX WPI; 1993-117468/14.
 XX Immuno-reactive hepatitis C virus polypeptide compsns. - contg.
 PT at least 2 sequences from the first variable domain of distinct
 PT HCV isolates
 XX Disclosure; Fig 3; 106pp; English.
 XX

CC The sequences given in AAR3392-002 represent a portion of the E2/NS1
 CC protein encoded by group I and group II HCV isolates, from amino acid
 CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
 CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
 CC 30 amino acids which shows large variation between nearly all isolates.
 CC This is an important immunoreactive domain. This putative envelope
 CC glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera
 CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
 CC flaviviruses, both of which confer protective immunity in hosts
 CC vaccinated with these polypeptides. It has been discovered that a
 CC number of important HCV epitopes vary among viral isolates and that
 CC these epitopes can be mapped to specific domains. This meant that
 CC immunologically cross-reactive polypeptides which focus on variable
 CC rather than constant domains can be produced. See also AAQ39134-48
 CC and AAR3392-91.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 144 AA;
 Query Match 52.4%; Score 11; DB 14; Length 144;
 Best Local Similarity 100.0%; Pred. No. 0.00052;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 LNCNESLNTGW 11
 |||||
 Db 58 LNCNESLNTGW 68
 |||||
 RESULT 9
 AAR34441
 ID AAR34441 standard; Protein; 174 AA.
 XX
 AC AAR34441;
 XX
 DT 25-MAR-2003 (updated)
 DT 09-AUG-1993 (first entry)
 XX
 DE Sequence of glycoprotein E2/NS1 in clone J1 (JM).
 XX
 KW Hepatitis C virus; envelope protein; glycoprotein; E2/NS1;
 KW diagnostic reagent.
 XX
 OS Hepatitis C virus.
 XX
 PN EP537626-A1.
 XX
 PD 21-APR-1993.
 XX
 PF 08-OCT-1992; 92EP-0117191.
 XX
 PR 08-OCT-1991; 91JP-0260824.
 XX
 PA (NAHE-) NAT INST OF HEALTH.
 XX
 PI Harada S, Honda Y, Miyamura T, Saito I;
 XX
 DR WPI; 1993-127516/16.
 DR N-PSDB; AAQ40333.
 XX
 PT Diagnostic reagent for hepatitis C virus - comprises second
 PT envelope protein or first non-structural protein encoded by HCV
 PT gene and has sugar chain
 XX
 PS Claim 2; Pages 40-41; 58pp; English.
 XX
 CC Glycoprotein E2/NS1 is derived from the second envelope protein or
 CC first non-structural protein encoded by the genome of HCV. The
 CC nucleic acid is extracted from the serum of the patient of hepatitis
 CC C. The serum is pref. mixed with transfer RNA (trna) as a carrier
 CC of virus RNA. As a technique of cloning cDNA from the nucleic acid,
 CC it is preferred to use polymerase chain reaction method. In the
 CC reaction, any commercially available random primers or synthesized
 CC DNA having a base sequence similar to that of primer AS1 may be used

CC as a primer. Representative examples of sense primers include S1.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 174 AA;
 Query Match 52.4%; Score 11; DB 14; Length 174;
 Best Local Similarity 100.0%; Pred. No. 0.00062;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 LNCNESLNTGW 11
 |||||
 Db 88 LNCNESLNTGW 98
 |||||
 RESULT 10
 AAU79220
 ID AAU79220 standard; Protein; 250 AA.
 XX
 AC AAU79220;
 XX
 DT 15-JUL-2002 (first entry)
 XX
 DE Hepatitis C Virus (HCV) delta-delta E2 genotype 1a protein.
 XX
 KW Hepatitis C Virus E2; HCV E2; virucide; hepatotropic; IL-2;
 KW antiinflammatory; HCV infection; interleukin-2; gamma-interferon;
 KW granulocyte macrophage-colony stimulating factor; GM-CSF;
 KW delta-delta E2 genotype 1a.
 XX
 OS Hepatitis C Virus.
 XX
 PN WO200222155-A1.
 XX
 PD 21-MAR-2002.
 XX
 PF 13-SEP-2001; 2001WO-US28767.
 XX
 PR 13-SEP-2000; 2000US-230927P.
 XX
 PA (HAWA-) HAWAII BIOTECHNOLOGY GROUP INC.
 XX
 PI Nakano ET, Clements DE, Humphreys T;
 XX
 DR WPI; 2002-383102/41.
 DR N-PSDB; ABK49386.
 XX
 PT New immunogenic polypeptide comprising hepatitis C virus E2 polypeptide
 PT useful for treating hepatitis C virus infection and for providing
 PT immune protection against virus infection -
 XX
 PS Claim 4; Fig 6; 84pp; English.
 XX
 CC The invention relates to a secreted polypeptide comprising hepatitis C
 CC virus (HCV) E2 polypeptide lacking all or a portion of its membrane
 CC spanning domain so that the E2 polypeptide is capable of secretion into
 CC growth medium when expressed recombinantly in a host cell. The
 CC polypeptide may also lack a portion of its C-terminus. The HCV E2
 CC secreted polypeptide is useful for producing anti-HCV antibodies. A
 CC purified immunogenic polypeptide comprising HCV E2 is useful for treating
 CC HCV infection and for providing immune protection against HCV infection
 CC by administering it to a subject having or at risk of having HCV
 CC infection or in need of protection. The method further comprises
 CC administering an immunomodulatory agent such as interleukin-2 (IL-2),
 CC granulocyte macrophage-colony stimulating factor (GM-CSF) or
 CC gamma-interferon. The polypeptide is useful as a vaccine, and with other
 CC HCV proteins to form a multi-component HCV vaccine for prophylactic or
 CC therapeutic treatment of HCV infection. This sequence represents an N-
 CC and C-terminally truncated HCV delta-delta E2 genotype 1a protein.
 XX
 SQ Sequence 250 AA;
 Query Match 52.4%; Score 11; DB 23; Length 250;
 Best Local Similarity 100.0%; Pred. No. 0.00085;

```
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 LNCNESLNTGW 11
    |||||
Db 16 LNCNESLNTGW 26

RESULT 11
AAE00445
ID AAE00445 standard; Protein; 271 AA.
AC AAE00445;
XX
XX 19-JUN-2001 (first entry)
XX
DE HCV E2-634 HVR1 construct containing E1 signal sequence and truncated E2.
XX
XX Hepatitis C virus; HCV; hypervariable region one; HVR1; vaccine;
KW antiviral; gene therapy; envelope 2 protein; E2; immunisation;
KW HCV infection; viral replication; passive immunoprophylaxis.
XX
XX Hepatitis C virus.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH Peptide 1..20
FT /label= Signal peptide
FT /note= "Derived from endoplasmic reticulum E1
FT signal sequence (364-383 amino acids)"
FT 21..271
FT /label= Mature_C_terminal_truncated_E2_protein
XX WO200121807-A1.
PN
XX 29-MAR-2001.
PD
XX 22-SEP-2000; 2000WO-US25987.
PF
XX 23-SEP-1999; 99US-0155823.
PR
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA
XX Forns X, Bukh J, Emerson SU, Purcell RH;
PI
XX WPI; 2001-266076/27.
XX
XX Novel nucleic acid molecules that encode hepatitis C virus envelope 2
PT protein lacking all or part of hypervariable region 1 of envelope
PT protein, useful as vaccine components for treating or preventing HCV
PT infections -
XX
XX Example; Page -; 80pp; English.
PS
XX The present sequence is hepatitis C virus (HCV) E2-634 HVR1 construct
CC containing the endoplasmic reticulum signal sequence of envelope protein
CC E1 and carboxy-terminal truncated E2 protein lacking the hypervariable
CC region one (HVR1). The HCV E2 protein lacking HVR1 DNA is useful for
CC producing infectious HCV and chimeric HCV viruses which are useful
CC for identifying cell lines capable of supporting the replication of
CC viruses. The infectious HCV and HVR1-chimeric HCV are used in the
CC production of attenuated or inactivated vaccines which are useful for
CC treating or preventing HCV in a mammal by immunisation. The host cells
CC expressing the H77C(HVR1) DNA is useful as an immunogen to stimulate a
CC protective immune response to HCV. The immunogens are useful for
CC producing protective antibodies to HCV. The antibodies produced are used
CC in passive immunoprophylaxis for treatment of diseases caused by HCV in
CC animals, especially humans. The H77C(HVR1) DNA is also useful in gene
CC therapy.
CC
CC Note: The present sequence is not shown in the specification but is
CC derived from Hepatitis C virus envelope 2 protein lacking hypervariable
CC region 1 (AAE00442) referred as SEQ ID NO: 2 and shown in figure 1.
XX
XX Sequence 271 AA;
```

```
Query Match 52.4%; Score 11; DB 22; Length 271;
Best Local Similarity 100.0%; Pred. No. 0.00091;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 LNCNESLNTGW 11
    |||||
Db 37 LNCNESLNTGW 47

RESULT 12
AAE00446
ID AAE00446 standard; Protein; 271 AA.
XX
XX AAE00446;
XX
XX 19-JUN-2001 (first entry)
XX
DE HCV E2-634 HVR1-mut containing E1 signal sequence and truncated E2.
XX
XX Hepatitis C virus; HCV; hypervariable region one; HVR1; vaccine;
KW antiviral; gene therapy; envelope 2 protein; E2; immunisation; mutant;
KW HCV infection; viral replication; passive immunoprophylaxis; mutant.
XX
XX Hepatitis C virus.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH Peptide 1..20
FT /label= Signal peptide
FT /note= "Derived from endoplasmic reticulum E1
FT signal sequence (364-383 amino acids)"
FT 21..271
FT /label= Mature_C_terminal_truncated_E2_protein
FT Misc-difference 225
FT /note= "Wild type Leu substituted with His"
XX WO200121807-A1.
PN
XX 29-MAR-2001.
PD
XX 22-SEP-2000; 2000WO-US25987.
PF
XX 23-SEP-1999; 99US-0155823.
PR
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA
XX Forns X, Bukh J, Emerson SU, Purcell RH;
PI
XX WPI; 2001-266076/27.
XX
XX Novel nucleic acid molecules that encode hepatitis C virus envelope 2
PT protein lacking all or part of hypervariable region 1 of envelope
PT protein, useful as vaccine components for treating or preventing HCV
PT infections -
XX
XX Example; Page -; 80pp; English.
PS
XX The present sequence is hepatitis C virus (HCV) E2-634 HVR1 construct
CC containing the endoplasmic reticulum signal sequence of envelope protein
CC E1 and carboxy-terminal truncated E2 protein lacking the hypervariable
CC region one (HVR1) with a replacement of leucine for histidine. The HCV
CC E2 protein lacking HVR1 DNA is useful for producing infectious HCV and
CC chimeric HCV viruses which are useful for identifying cell lines
CC capable of supporting the replication of viruses. The infectious HCV and
CC HVR1-chimeric HCV are used in the production of attenuated or inactivated
CC vaccines which are useful for treating or preventing HCV in a mammal by
CC immunisation. The host cells expressing the H77C(HVR1) DNA is useful as
CC an immunogen to stimulate a protective immune response to HCV. The
CC immunogens are useful for producing protective antibodies to HCV. The
CC antibodies produced are used in passive immunoprophylaxis for treatment
CC of diseases caused by HCV in animals, especially humans. The H77C(HVR1)
CC DNA is also useful in gene therapy.
```

CC Note: The present sequence is not shown in the specification but is
CC derived from Hepatitis C virus envelope 2 protein lacking hypervariable
CC region 1 (AAE00442) referred as SEQ ID NO: 2 and shown in figure 1.

XX Sequence 271 AA;

Query Match 52.4%; Score 11; DB 22; Length 271;
Best Local Similarity 100.0%; Pred. No. 0.00091;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
| | | | | | | | | |
Db 37 LNCNESLNTGW 47

RESULT 13
AAW00929
ID AAW00929 standard; Protein; 305 AA.

XX AC

XX AAW00929;

XX DT 04-NOV-1997 (first entry)

XX DE Recombinant HCV E2 antigen.

XX KW HCV; E2; antigen; non-secretor gene; protein secretion; vaccine;

XX OS Chimeric hepatitis C virus;

XX OS Chimeric synthetic.

XX FH Key Location/Qualifiers

FT Peptide 1..19

FT Cleavage-site 20 /label= Sig_peptide

FT /note= "mammalian secretion signal

FT /label= Mat_protein

FT Peptide 25..28

FT /note= "N-terminal sequence of prourokinase"

XX WO9641179-A1.

XX PD 19-DEC-1996.

XX PF 05-JUN-1996; 96WO-US09345.

XX PR 07-JUN-1995; 95US-0478073.

XX PA (ABBO) ABBOTT LAB.

XX PI Lesniewski RR, Okasinski GF, Schaefer VG, Suhar TS;

XX DR WPI; 1997-108653/10.

XX N-PSDB; AAT13899.

XX PT New expression system for proteins, partic. HCV antigens - for use

PT in assays for screening and prognostic applications and for use in

PT vaccines

XX Example 1; Page 29-30; 40pp; English.

XX PS

CC This sequence comprises the conceptual translation product of

CC a hepatitis C virus E2 antigen expression cassette (AAT13899). The

CC encoded E2 antigen is truncated at amino acid residue 644 of HCV,

CC and contains an N-terminal sequence (SNEI) from human prourokinase

CC intended to promote signal processing, efficient secretion

CC and final product stability. Glycosylated E2 can be expressed in

CC mammalian host cells utilizing claimed plasmid 577. Fusion

CC proteins produced by plasmid 577 are used in claimed assays and

CC test kits for detecting anti-analyte antibody as well as in claimed

CC vaccines for treatment of infection. The plasmid can be used to

CC produce high levels of proteins that would not normally be
CC expressed in mammalian cells due to the non-secretory nature of the
CC gene. The system is esp. used for expressing HCV proteins, allowing
CC proper processing, glycosylation and conformation of the viral
CC protein.

XX SQ Sequence 305 AA;

Query Match 52.4%; Score 11; DB 18; Length 305;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
| | | | | | | | | |
Db 68 LNCNESLNTGW 78

RESULT 14
AAR45330

ID AAR45330 standard; protein; 319 AA.

XX AC AAR45330;

XX DT 28-JUN-1994 (first entry)

XX DE Anti-HCV antibody reactive protein #1.

XX KW Hepatitis C virus; HCV; envelope; region; anti-HCV; antibody; vaccine.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Peptide 39..69

FT /label= Epi

FT /note= "Claim 1"

XX JP05310786-A.

XX PD 22-NOV-1993.

XX PF 30-APR-1992; 92JP-0111846.

XX PR 30-APR-1992; 92JP-0111846.

XX PA (OLYU) OLYMPUS OPTICAL CO LTD.

XX DR WPI; 1994-002315/01.

XX PT Peptide of HCV envelope region - having high reactivity against

PT anti-hepatitis C virus antibody and having at least 6 amino acids

XX Disclosure; Page 6-7; 11pp; Japanese.

XX The sequences given in AAR45330-31 represent fragments derived from
CC the hepatitis C virus (HCV) envelope region that are highly reactive
CC against anti-HCV antibodies. These protein fragments may be used in
CC vaccines against HCV. They have a high reactivity against anti-HCV
CC antibodies regardless of the original HCV. They can be used to
CC detect HCV infection but not the onset of infection.

XX SQ Sequence 319 AA;

Query Match 52.4%; Score 11; DB 15; Length 319;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
| | | | | | | | | |
Db 236 LNCNESLNTGW 246

RESULT 15
AAR40118

```

ID  AAR40118 standard; Protein; 333 AA.
XX
AC  AAR40118;
XX
DT  25-MAR-2003 (updated)
DT  27-JAN-1994 (first entry)
XX
DE  HGH-HCV-E2 fusion protein expressed by pHCV-170.
XX
KW  Hepatitis C Virus; Non-A, non-B hepatitis Virus; HCV; NANBHV;
KW  human growth hormone; HGH; secretion signal; fusion protein;
KW  vaccine.
XX
OS  Chimeric Hepatitis C Virus.
OS  Chimeric Homo sapiens.
XX
FH  Key Location/Qualifiers
FT  Region 1..32
FT  /note= "HGH secretion signal"
FT  Region 33..333
FT  /note= "HCV-E2 (amino acids 384-684)"
XX
PN  WO9315193-A1.
XX
PD  05-AUG-1993.
XX
PF  29-JAN-1993; 93WO-US00907.
XX
PR  31-JAN-1992; 92US-0830024.
XX
PA  (ABBO ) ABBOTT LAB.
XX
PI  Bode SL, Casey JM, Desai SM, Devare SG, Frail DE;
PI  Yamaguchi J, Zeck BJ;
XX
DR  WPI; 1993-258673/32.
DR  N-PSDB; AAQ47195.
XX
XX  New plasmid pHCV-162 is a mammalian expression systems for HCV
PT  proteins - useful for diagnosing HCV infection and as vaccines
PT  for preventing HCV infection
XX
PS  Claim 10; Page 81-82; 100pp; English.
XX
CC  A sequence coding for the HCV E2 protein (amino acids 384-684 from
CC  HCV-infected chimpanzee CO isolate, see AAR40119) was generated using
CC  PCR. An EcoRI site was used to attach a synthetic oligonucleotide
CC  encoding the Human Growth Hormone secretion signal at the 5'-end of
CC  the HCV sequence. The resultant plasmid was designated pHCV-170.
CC  AAR40118 is the sequence of the HGH-HCV-E2 fusion protein expressed
CC  by vector pHCV-170. See AAQ47192-Q47196.
CC  (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ  Sequence 333 AA;

Query Match 52.4%; Score 11; DB 14; Length 333;
Best Local Similarity 100.0%; Pred.No. 0.0011;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LNCNESLNTGW 11
   |||||
DB 76 LNCNESLNTGW 86

Search completed: November 21, 2003, 20:58:02
Job time : 33.755 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 11.13 Seconds
(without alignments)
79.832 Million cell updates/sec

Title: US-09-973-025-74
Perfect score: 21
Sequence: 1 LNCNLSLNTGWLGLIYQHK 21

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0
Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	21	US-08-612-973-74	Sequence 74, Appl
2	21	100.0	21	US-08-927-597-74	Sequence 74, Appl
3	21	100.0	21	US-08-635-886C-30	Sequence 30, Appl
4	11	52.4	38	US-08-262-037-93	Sequence 93, Appl
5	11	52.4	40	US-08-262-037-10	Sequence 10, Appl
6	11	52.4	52	US-08-262-037-94	Sequence 94, Appl
7	11	52.4	144	US-08-440-103-20	Sequence 20, Appl
8	11	52.4	144	US-08-440-542-20	Sequence 20, Appl
9	11	52.4	144	US-08-231-368-20	Sequence 20, Appl
10	11	52.4	144	US-08-440-210-20	Sequence 20, Appl
11	11	52.4	144	US-09-046-604-20	Sequence 20, Appl
12	11	52.4	174	US-08-460-806-19	Sequence 19, Appl
13	11	52.4	174	US-08-325-630-19	Sequence 19, Appl
14	11	52.4	180	US-08-483-695-40	Sequence 40, Appl
15	11	52.4	180	US-07-965-285-40	Sequence 40, Appl
16	11	52.4	180	US-08-487-231-40	Sequence 40, Appl
17	11	52.4	180	US-09-201-912-40	Sequence 40, Appl
18	11	52.4	305	US-08-478-073-2	Sequence 2, Appl
19	11	52.4	333	US-08-453-552-12	Sequence 12, Appl
20	11	52.4	333	US-08-710-637-12	Sequence 12, Appl
21	11	52.4	333	PCT-US93-00907-12	Sequence 12, Appl
22	11	52.4	337	US-08-188-281B-7	Sequence 7, Appl
23	11	52.4	337	PCT-US94-07280-7	Sequence 7, Appl
24	11	52.4	337	PCT-US95-01087-7	Sequence 7, Appl
25	11	52.4	367	US-08-188-281B-9	Sequence 9, Appl
26	11	52.4	367	US-08-453-552-6	Sequence 6, Appl
27	11	52.4	367	US-08-710-637-6	Sequence 6, Appl

28	11	52.4	367	5	PCT-US93-00907-6	Sequence 6, Appl
29	11	52.4	367	5	PCT-US94-07280-9	Sequence 9, Appl
30	11	52.4	367	5	PCT-US95-01087-9	Sequence 9, Appl
31	11	52.4	377	1	US-08-188-281B-17	Sequence 17, Appl
32	11	52.4	377	5	PCT-US94-07280-17	Sequence 17, Appl
33	11	52.4	377	5	PCT-US95-01087-17	Sequence 17, Appl
34	11	52.4	397	1	US-08-188-281B-11	Sequence 11, Appl
35	11	52.4	397	5	PCT-US94-07280-11	Sequence 11, Appl
36	11	52.4	397	5	PCT-US95-01087-11	Sequence 11, Appl
37	11	52.4	399	1	US-08-453-552-10	Sequence 10, Appl
38	11	52.4	399	2	US-08-710-637-10	Sequence 10, Appl
39	11	52.4	399	5	PCT-US93-00907-10	Sequence 10, Appl
40	11	52.4	402	1	US-08-460-806-15	Sequence 15, Appl
41	11	52.4	402	1	US-08-325-630-15	Sequence 15, Appl
42	11	52.4	409	1	US-08-440-103-21	Sequence 21, Appl
43	11	52.4	409	1	US-08-440-103-24	Sequence 24, Appl
44	11	52.4	409	1	US-08-440-542-21	Sequence 21, Appl
45	11	52.4	409	1	US-08-440-542-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-08-612-973-74
; Sequence 74, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/POCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-612-973-74

Query Match 100.0%; Score 21; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 1e-14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LNCNLSLNTGWLGLIYQHK 21
|||||

Db 1 LNCNLSLNTGWLWLAGLIYQHK 21

RESULT 2

US-08-927-597-74

Sequence 74, Application US/08927597

Patent No. 6245503

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

APPLICANT: BOSMAN, FONS

APPLICANT: DE MARTYNOFF, GUY

APPLICANT: BUYSSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/927,597

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 74:

SEQUENCE CHARACTERISTICS:

LENGTH: 21 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-927-597-74

Query Match 100.0%; Score 21; DB 3; Length 21;

Best Local Similarity 100.0%; Pred. No. 1e-14;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNLSLNTGWLWLAGLIYQHK 21

Db 1 LNCNLSLNTGWLWLAGLIYQHK 21

RESULT 3

US-08-635-886C-30

Sequence 30, Application US/08635886C

Patent No. 6555114

GENERAL INFORMATION:

APPLICANT: LEROUX-ROELS, Geert

APPLICANT: DELEYS, Robert

APPLICANT: MAERTENS, Geert

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C

TITLE OF INVENTION: VIRUS

FILE REFERENCE: 2752-18

CURRENT APPLICATION NUMBER: US/08/635,886C

CURRENT FILING DATE: 1996-04-25

PRIOR APPLICATION NUMBER: PCT/EP94/03555

PRIOR FILING DATE: 1994-10-28

PRIOR APPLICATION NUMBER: EP 93402718.6

PRIOR FILING DATE: 1993-11-04

NUMBER OF SEQ ID NOS: 286

SOFTWARE: PatentIn version 3.1

SEQ ID NO 30

LENGTH: 21

TYPE: PRT

ORGANISM: hepatitis C virus

US-08-635-886C-30

Query Match 100.0%; Score 21; DB 4; Length 21;

Best Local Similarity 100.0%; Pred. No. 1e-14;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNLSLNTGWLWLAGLIYQHK 21

Db 1 LNCNLSLNTGWLWLAGLIYQHK 21

RESULT 4

US-08-262-037-93

Sequence 93, Application US/08262037

Patent No. 5747239

GENERAL INFORMATION:

APPLICANT: Chang Yi Wang and Barbara Hosein

TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR

TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV

TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES

NUMBER OF SEQUENCES: 136

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVE.

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/262,037

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/719,819

FILING DATE: 24-June-1991

APPLICATION NUMBER: 07/667,275

FILING DATE: 11-Mar-1991

APPLICATION NUMBER: 07/651,735

FILING DATE: 07-Feb-1991

APPLICATION NUMBER: 07/558,799

FILING DATE: 26-July-1990

APPLICATION NUMBER: 07/510,153

FILING DATE: 16-April-1990

ATTORNEY/AGENT INFORMATION:

NAME: Maria C. H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4043 US3

TELEPHONE: 212-758-4800

TELEFAX: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 93:

SEQUENCE CHARACTERISTICS:

LENGTH: 38 amino acids

TYPE: Amino acid

STRANDEDNESS:

TOPOLOGY: Unknown

US-08-262-037-93

STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,103
FILING DATE: 12-MAY-1995
CLASSIFICATION: 424
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-103-20

```
Query Match      52.4%; Score 11; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 0.00042;
Matches 11; Conservative 0; Mismatches 0; Indels
```

Qy 1 LNCNESLNTGW 11
Db 58 LNCNESLNTGW 68

```

RESULT 8
US-08-440-542-20
; Sequence 20, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,542
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 20:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-542-20

```

Query Match 52.4%; Score 11; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 0.00042;
Matches 11; Conservative 0; Mismatches 0; Indels

Qy 1 LNCNESLNTGW 11
Db 58 LNCNESLNTGW 68

RESULT 9
 US-08-231-368-20
 ; Sequence 20, Application US/08231368
 ; Patent No. 5756312
 ; GENERAL INFORMATION:
 ; APPLICANT: Weiner, Amy J.
 ; APPLICANT: Houghton, Michael
 ; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
 ; NUMBER OF SEQUENCES: 45
 ; CORRESPONDENCE ADDRESS:

Query Match 52.4%; Score 11; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 0.00042;
Matches 11: Conservative 0; Mismatches 0; Indels

Ov 1 LNCNESLNTGW 11

```
Db          ||||| 58 LNCNESLNTGW 68
;
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/046,604
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-046-604-20

Query Match          52.4%; Score 11; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 0.00042;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          1 LNCNESLNTGW 11
Db          ||||| 58 LNCNESLNTGW 68
;
; RESULT 12
; US-08-460-806-19
; Sequence 19, Application US/08460806
; Patent No. 5747241
; GENERAL INFORMATION:
; APPLICANT: MIYAMURA, TATSUO
; APPLICANT: SAITO, IZUMU
; APPLICANT: HARADA, SHIZUKO
; APPLICANT: HONDA, YOSHIKAZU
; TITLE OF INVENTION: DIAGNOSTIC REAGENT FOR HEPATITIS C
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,806
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,630
; FILING DATE: 19-OCT-1994
; APPLICATION NUMBER: US 07/956,993

;
; RESULT 11
; US-09-046-604-20
; Sequence 20, Application US/09046604
; Patent No. 6303292
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA

;
; RESULT 10
; US-08-440-210-20
; Sequence 20, Application US/08440210
; Patent No. 5766845
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,210
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-210-20

Query Match          52.4%; Score 11; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 0.00042;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          1 LNCNESLNTGW 11
Db          ||||| 58 LNCNESLNTGW 68
;
; RESULT 11
; US-09-046-604-20
; Sequence 20, Application US/09046604
; Patent No. 6303292
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
```

FILING DATE: 06-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5747241man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 4667-001-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 174 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-460-806-19

Query Match 52.4%; Score 11; DB 1; Length 174;
Best Local Similarity 100.0%; Pred. No. 0.0005;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
Db 88 LNCNESLNTGW 98

RESULT 13
US-08-325-630-19
Sequence 19, Application US/08325630
Patent No. 5750331
GENERAL INFORMATION:
APPLICANT: MIYAMURA, TATSUO
APPLICANT: SAITO, IZUMU
APPLICANT: HARADA, SHIZUKO
APPLICANT: HONDA, YOSHIKAZU
TITLE OF INVENTION: DIAGNOSTIC REAGENT FOR HEPATITIS C
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325,630
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/956,993
FILING DATE: 06-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5750331man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 4667-001-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 174 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-325-630-19

Query Match 52.4%; Score 11; DB 1; Length 174;
Best Local Similarity 100.0%; Pred. No. 0.0005;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
Db 88 LNCNESLNTGW 98

RESULT 14
US-08-483-695-40
Sequence 40, Application US/08483695
Patent No. 5866139
GENERAL INFORMATION:
APPLICANT: Brechot, Christian
APPLICANT: Krensdorff, Dina
APPLICANT: Porchon, Colette
TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
TITLE OF INVENTION: Applications
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3315

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,695
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/965,285
FILING DATE: 18-MAR-1993
APPLICATION NUMBER: FR 91 06 882
FILING DATE: 06-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 05286-0001-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400

INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 180 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-483-695-40

Query Match 52.4%; Score 11; DB 2; Length 180;
Best Local Similarity 100.0%; Pred. No. 0.00052;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
Db 94 LNCNESLNTGW 104

RESULT 15
US-07-965-285-40
Sequence 40, Application US/07965285
Patent No. 5879904
GENERAL INFORMATION:
APPLICANT: Brechot, Christian

APPLICANT: Kremsdorf, Dina
APPLICANT: Porchon, Colette
TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
TITLE OF INVENTION: Applications
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/965,285
FILING DATE: 18-MAR-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 06 882
FILING DATE: 06-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 05286-0001-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 180 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-965-285-40

Query Match 52.4%; Score 11; DB 2; Length 180;
Best Local Similarity 100.0%; Pred. No. 0.00052;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNLSLNTGW 11
Db 94 LNCNLSLNTGW 104

Search completed: November 21, 2003, 21:15:14
Job time : 11.13 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time 20.475 Seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-74

Perfect score: 21

Sequence: 1 LNCNLSNTGWLGLIYQHK 21

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 18259486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	21	10	US-09-973-025-74
2	21	100.0	21	11	US-09-973-025-74
3	21	100.0	21	11	US-09-999-303-74
4	21	100.0	21	11	US-09-995-808-74
5	21	100.0	21	12	US-09-995-860-74
6	11	52.4	91	10	US-09-995-791-74
7	11	52.4	250	10	US-09-921-397-80
8	11	52.4	350	10	US-09-952-572-8
9	11	52.4	350	14	US-09-929-955-4
10	11	52.4	363	12	US-10-104-966-4
11	11	52.4	363	15	US-10-128-587A-97
12	11	52.4	2894	10	US-09-941-611-23
13	11	52.4	2894	15	US-10-044-995-23
14	11	52.4	3011	9	US-09-742-659-4
15	11	52.4	3011	10	US-09-952-572-9

16	11	52.4	3011	10	US-09-929-955-1	Sequence 1, Appli
17	11	52.4	3011	10	US-09-747-419-20	Sequence 20, Appli
18	11	52.4	3011	11	US-09-891-894-3	Sequence 3, Appli
19	11	52.4	3011	12	US-10-184-150-3	Sequence 3, Appli
20	11	52.4	3011	14	US-10-104-966-1	Sequence 1, Appli
21	11	52.4	3011	15	US-10-259-275-20	Sequence 20, Appli
22	11	52.4	3012	10	US-09-238-076-2	Sequence 2, Appli
23	11	52.4	3012	11	US-09-995-937-2	Sequence 2, Appli
24	11	52.4	3012	11	US-09-917-563-2	Sequence 2, Appli
25	08	38.1	20	10	US-09-973-025-75	Sequence 75, Appli
26	8	38.1	20	11	US-09-899-303-75	Sequence 75, Appli
27	8	38.1	20	11	US-09-995-808-75	Sequence 75, Appli
28	8	38.1	20	11	US-09-995-860-75	Sequence 75, Appli
29	8	38.1	20	12	US-09-995-791-75	Sequence 75, Appli
30	6	28.6	130	15	US-10-097-065-230	Sequence 230, App
31	6	28.6	195	12	US-10-029-386-34010	Sequence 34010, A
32	6	28.6	241	12	US-10-193-109-8	Sequence 8, Appli
33	6	28.6	241	15	US-10-084-994-8	Sequence 8, Appli
34	6	28.6	249	15	US-10-156-761-14492	Sequence 14492, A
35	6	28.6	254	10	US-09-407-430-3	Sequence 3, Appli
36	6	28.6	347	11	US-09-194-949-9	Sequence 9, Appli
37	6	28.6	363	10	US-09-407-430-2	Sequence 2, Appli
38	6	28.6	384	15	US-10-106-698-6253	Sequence 6253, Ap
39	6	28.6	390	12	US-09-944-049-20	Sequence 20, Appli
40	6	28.6	412	12	US-09-944-049-22	Sequence 22, Appli
41	6	28.6	539	11	US-09-194-949-11	Sequence 11, Appli
42	6	28.6	637	12	US-10-187-257-4	Sequence 2, Appli
43	6	28.6	637	12	US-10-265-083-2	Sequence 2, Appli
44	6	28.6	900	10	US-09-738-626-6108	Sequence 6108, Ap
45	6	28.6	2202	12	US-10-094-466-12	Sequence 12, Appli

ALIGNMENTS

RESULT 1

US-09-973-025-74
; Sequence 74, Application US/09973025
; Publication No. US20020182706A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973.025
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100

; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-09-973-025-74

Query Match 100.0%; Score 21; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.2e-15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGWLGLIYQHK 21
| | | | | | | | | | | | | | | | | | | | |
Db 1 LNCNESLNTGWLGLIYQHK 21

RESULT 2
US-09-899-303-74
; Sequence 74, Application US/09899303
; Publication No. US2003036110A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-09-899-303-74

Query Match 100.0%; Score 21; DB 11; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.2e-15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGWLGLIYQHK 21
| | | | | | | | | | | | | | | | | | | | |

Db 1 LNCNESLNTGWLGLIYQHK 21

RESULT 3
US-09-995-808-74
; Sequence 74, Application US/09995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent in 3.1
; SEQ ID NO 74
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-74

Query Match 100.0%; Score 21; DB 11; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.2e-15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGWLGLIYQHK 21
| | | | | | | | | | | | | | | | | | | | |
Db 1 LNCNESLNTGWLGLIYQHK 21

RESULT 4
US-09-995-860-74
; Sequence 74, Application US/09995860
; Publication No. US20030118603A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent in 3.1
; SEQ ID NO 74
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-74

Query Match 100.0%; Score 21; DB 11; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.2e-15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGWLGLIYQHK 21
| | | | | | | | | | | | | | | | | | | | |
Db 1 LNCNESLNTGWLGLIYQHK 21

RESULT 5
US-09-995-791-74
; Sequence 74, Application US/09995791
; Publication No. US20030147918A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent in 3.1
; SEQ ID NO 74

Query Match 100.0%; Score 21; DB 11; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.2e-15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; LENGTH: 21
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-74

Query Match 100.0%; Score 21; DB 12; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.2e-15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGWLGLIYQHK 21
Db 1 LNCNESLNTGWLGLIYQHK 21

RESULT 6

US-09-921-397-80
; Sequence 80, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 80
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-80

Query Match 52.4%; Score 11; DB 10; Length 91;
Best Local Similarity 100.0%; Pred. No. 0.00028;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
Db 48 LNCNESLNTGW 58

RESULT 7

US-09-952-572-8
; Sequence 8, Application US/09952572
; Patent No. US20020119495A1
; GENERAL INFORMATION:
; APPLICANT: HAWAII BIOTECHNOLOGY GROUP, Inc.
; APPLICANT: NAKANO, Eileen
; APPLICANT: CLEMENTS, David
; APPLICANT: HUMPHREYS, Tom
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITION OF HEPATITIS C
; FILE REFERENCE: HAWBIO1100
; CURRENT APPLICATION NUMBER: US/09/952,572
; CURRENT FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: US 60/230,927
; PRIOR FILING DATE: 2000-09-13
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-09-952-572-8

Query Match 52.4%; Score 11; DB 10; Length 250;
Best Local Similarity 100.0%; Pred. No. 0.00066;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11

Db 16 LNCNESLNTGW 26
|||||

RESULT 8

US-09-929-955-4
; Sequence 4, Application US/09929955
; Patent No. US20020136740A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; FILE REFERENCE: TRIPEP.23AUS2
; CURRENT APPLICATION NUMBER: US/09/929,955
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/225,767
; PRIOR FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-09-929-955-4

Query Match 52.4%; Score 11; DB 10; Length 350;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
Db 48 LNCNESLNTGW 58
|||||

RESULT 9

US-10-104-966-4
; Sequence 4, Application US/10104966
; Publication No. US20020155124A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; FILE REFERENCE: TRIPEP.23AUSC1
; CURRENT APPLICATION NUMBER: US/10/104,966
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-10-104-966-4

Query Match 52.4%; Score 11; DB 14; Length 350;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
|||||

Db 48 LNCNESLNTGW 58

RESULT 10
US-10-128-587A-97
; Sequence 97, Application US/10128587A
; Publication No. US20030152940A1
; GENERAL INFORMATION:
; APPLICANT: Imogenetics N.V.
; TITLE OF INVENTION: Constructs and methods for expression of recombinant HCV envelope
; FILE REFERENCE: 134 PCT
; CURRENT APPLICATION NUMBER: US/10/128,587A
; CURRENT FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 97
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: hepatitis C virus
US-10-128-587A-97

Query Match 52.4%; Score 11; DB 12; Length 363;
Best Local Similarity 100.0%; Pred. No. 0.0009;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
|||||
Db 44 LNCNESLNTGW 54

RESULT 11
US-10-128-590-97
; Sequence 97, Application US/10128590
; Publication No. US20030108561A1
; GENERAL INFORMATION:
; APPLICANT: Imogenetics N.V.
; TITLE OF INVENTION: Core-glycosylated HCV envelope proteins
; FILE REFERENCE: 135 PCT
; CURRENT APPLICATION NUMBER: US/10/128,590
; CURRENT FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 97
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-128-590-97

Query Match 52.4%; Score 11; DB 15; Length 363;
Best Local Similarity 100.0%; Pred. No. 0.0009;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
|||||
Db 44 LNCNESLNTGW 54

RESULT 12
US-09-941-611-23
; Sequence 23, Application US/09941611
; Patent No. US20020106640A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; POLLET, DIRK
; MAERTENS, GEERT
; VAN HEUVERSWIN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; ANTIBODIES TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/09/941,611
FILING DATE: 30-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: 1995-02-21
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
o LENGTH: 2894 amino acids
o TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-941-611-23

Query Match 52.4%; Score 11; DB 10; Length 2894;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
|||||
Db 427 LNCNESLNTGW 437

RESULT 13
US-10-044-995-23
; Sequence 23, Application US/10044995
; Publication No. US20030049685A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; POLLET, DIRK
; MAERTENS, GEERT
; VAN HEUVERSWIN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; ANTIBODIES TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
o MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/044,995
FILING DATE: 15-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: <Unknown>
APPLICATION NUMBER: US 07/920,286
FILING DATE: 14-OCT-1992
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-044-995-23

Query Match 52.4%; Score 11; DB 15; Length 2894;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
Db 427 LNCNESLNTGW 437

RESULT 14
US-09-742-659-4
Sequence 4, Application US/09742659
Patent No. US20010034019A1
GENERAL INFORMATION:
APPLICANT: Hong, Zhi
APPLICANT: Butkiewicz, Nancy J.
APPLICANT: Zhong, Weidong
APPLICANT: Ingravallo, Paul
APPLICANT: Wright-Minoque, Jacquelyn
APPLICANT: Lau, Johnson Y.
APPLICANT: Lemon, Stanley M.
TITLE OF INVENTION: Chimeric HCV/GBV-B viruses
FILE REFERENCE: ID0116
CURRENT APPLICATION NUMBER: US/09/742,659
CURRENT FILING DATE: 2000-12-21
PRIOR APPLICATION NUMBER: US 60/171,469
PRIOR FILING DATE: 1999-12-22
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 3011
TYPE: PRT
ORGANISM: Hepatitis C virus
US-09-742-659-4

Query Match 52.4%; Score 11; DB 9; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11

Db 427 LNCNESLNTGW 437
RESULT 15
US-09-952-572-9
Sequence 9, Application US/09952572
Patent No. US20020119495A1
GENERAL INFORMATION:
APPLICANT: HAWAII BIOTECHNOLOGY GROUP, Inc.
APPLICANT: NAKANO, Eileen
APPLICANT: CLEMENTS, David
APPLICANT: HUMPHREYS, Tom
TITLE OF INVENTION: IMMUNOGENIC COMPOSITION OF HEPATITIS C
FILE REFERENCE: HAWBIO1100
CURRENT APPLICATION NUMBER: US/09/952,572
CURRENT FILING DATE: 2001-09-13
PRIOR APPLICATION NUMBER: US 60/230,927
PRIOR FILING DATE: 2000-09-13
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn version 3.0
SEQ ID NO 9
LENGTH: 3011
TYPE: PRT
ORGANISM: Hepatitis C Virus
US-09-952-572-9

Query Match 52.4%; Score 11; DB 10; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
Db 427 LNCNESLNTGW 437

Search completed: November 21, 2003, 22:19:37
Job time : 21.475 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 171.255 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-74

Perfect score: 21

Sequence: 1 LNCNLSMTGWLGLIYQHK 21

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 309918778 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents AA Main.*
1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US097A_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US097B_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US099A_COMB.pep.*
25: /cgn2_6/ptodata/1/paa/US099B_COMB.pep.*
26: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
27: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
28: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
29: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
30: /cgn2_6/ptodata/1/paa/US104_COMB.pep.*
31: /cgn2_6/ptodata/1/paa/US106_COMB.pep.*
32: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	21	13	US-08-974-690C-30
					Sequence 30, Appl

ALIGNMENTS

RESULT 1
US-08-974-690C-30
; Sequence 30, Application US/08974690C
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 21
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-974-690C-30

2	21	100.0	21	23	US-09-899-303-74	Sequence 74, Appl
3	21	100.0	21	23	US-09-899-303A-74	Sequence 74, Appl
4	21	100.0	21	25	US-09-973-025-74	Sequence 74, Appl
5	21	100.0	21	25	US-09-995-791-74	Sequence 74, Appl
6	21	100.0	21	25	US-09-995-808-74	Sequence 74, Appl
7	21	100.0	21	25	US-09-995-860-74	Sequence 74, Appl
8	21	100.0	21	26	US-10-020-510-74	Sequence 74, Appl
9	21	100.0	21	29	US-10-321-798-74	Sequence 74, Appl
10	11	52.4	20	13	US-08-974-685-30	Sequence 30, Appl
11	11	52.4	20	13	US-08-974-690-30	Sequence 30, Appl
12	11	52.4	20	13	US-08-974-690A-30	Sequence 30, Appl
13	11	52.4	20	13	US-08-974-690B-30	Sequence 30, Appl
14	11	52.4	27	1	PCT-US02-17374-24	Sequence 24, Appl
15	11	52.4	38	8	US-08-475-482-93	Sequence 93, Appl
16	11	52.4	38	8	US-08-477-072-93	Sequence 93, Appl
17	11	52.4	38	8	US-08-477-582-93	Sequence 93, Appl
18	11	52.4	38	8	US-08-480-253-93	Sequence 93, Appl
19	11	52.4	40	3	US-07-667-275A-10	Sequence 10, Appl
20	11	52.4	40	8	US-08-475-482-10	Sequence 10, Appl
21	11	52.4	40	8	US-08-477-072-10	Sequence 10, Appl
22	11	52.4	40	8	US-08-477-582-10	Sequence 10, Appl
23	11	52.4	40	8	US-08-480-253-10	Sequence 10, Appl
24	11	52.4	52	8	US-08-475-482-94	Sequence 94, Appl
25	11	52.4	52	8	US-08-477-072-94	Sequence 94, Appl
26	11	52.4	52	8	US-08-477-582-94	Sequence 94, Appl
27	11	52.4	52	8	US-08-480-253-94	Sequence 94, Appl
28	11	52.4	91	24	US-09-921-397-80	Sequence 80, Appl
29	11	52.4	144	8	US-08-471-498-20	Sequence 20, Appl
30	11	52.4	250	1	PCT-US01-28767-8	Sequence 8, Appl
31	11	52.4	250	25	US-09-952-572-8	Sequence 8, Appl
32	11	52.4	333	8	US-08-417-478-12	Sequence 12, Appl
33	11	52.4	333	8	US-08-453-613-12	Sequence 12, Appl
34	11	52.4	350	21	US-09-705-547-4	Sequence 4, Appl
35	11	52.4	350	24	US-09-929-955-4	Sequence 4, Appl
36	11	52.4	350	27	US-10-104-966-4	Sequence 4, Appl
37	11	52.4	350	32	US-60-229-175-4	Sequence 4, Appl
38	11	52.4	363	27	US-10-128-587A-97	Sequence 97, Appl
39	11	52.4	363	27	US-10-128-590-97	Sequence 97, Appl
40	11	52.4	367	8	US-08-417-478-6	Sequence 6, Appl
41	11	52.4	367	8	US-08-453-613-6	Sequence 6, Appl
42	11	52.4	399	8	US-08-417-478-10	Sequence 10, Appl
43	11	52.4	399	8	US-08-453-613-10	Sequence 10, Appl
44	11	52.4	409	8	US-08-471-498-21	Sequence 21, Appl
45	11	52.4	409	8	US-08-471-498-24	Sequence 24, Appl

Query Match 100.0%; Score 21; DB 13; Length 21;
Best Local Similarity 100.0%; Pred. No. 6.8e-15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGWLGLIYQHK 21
| | | | | | | | | | | | | | | | | | | | |
Db 1 LNCNESLNTGWLGLIYQHK 21

RESULT 2

US-09-899-303-74
; Sequence 74, Application US/09899303
; GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/899,303

FILING DATE: 06-Jul-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 74:

SEQUENCE CHARACTERISTICS:

LENGTH: 21 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 74:

US-09-899-303-74

Query Match 100.0%; Score 21; DB 23; Length 21;
Best Local Similarity 100.0%; Pred. No. 6.8e-15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGWLGLIYQHK 21
| | | | | | | | | | | | | | | | | | | | |
Db 1 LNCNESLNTGWLGLIYQHK 21

RESULT 3

US-09-899-303A-74

; Sequence 74, Application US/09899303A

; GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

BOSMAN, FONS

DE MARTYNOFF, GUY

BUYSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/899,303A

FILING DATE: 06-Jul-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 74:

SEQUENCE CHARACTERISTICS:

LENGTH: 21 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 74:

US-09-899-303A-74

Query Match 100.0%; Score 21; DB 23; Length 21;
Best Local Similarity 100.0%; Pred. No. 6.8e-15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGWLGLIYQHK 21
| | | | | | | | | | | | | | | | | | | | |
Db 1 LNCNESLNTGWLGLIYQHK 21

RESULT 4

US-09-973-025-74

; Sequence 74, Application US/09973025

; GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

BOSMAN, FONS

DE MARTYNOFF, GUY

BUYSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/973,025

; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-09-973-025-74

Query Match 100.0%; Score 21; DB 25; Length 21;
Best Local Similarity 100.0%; Pred. No. 6.8e-15; Mismatches 0; Indels 0; Gaps 0;
Matches 21; Conservative 0;

Qy 1 LNCNESLNTGWWLAGLIYQHK 21
| | | | | | | | | | | | | | | | | | | | |
Db 1 LNCNESLNTGWWLAGLIYQHK 21

RESULT 5
US-09-995-791-74
; Sequence 74, Application US/09995791
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 74
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-74

Query Match 100.0%; Score 21; DB 25; Length 21;
Best Local Similarity 100.0%; Pred. No. 6.8e-15; Mismatches 0; Indels 0; Gaps 0;
Matches 21; Conservative 0;

Qy 1 LNCNESLNTGWWLAGLIYQHK 21
| | | | | | | | | | | | | | | | | | | | |
Db 1 LNCNESLNTGWWLAGLIYQHK 21

RESULT 6
US-09-995-808-74
; Sequence 74, Application US/09995808
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 74
; LENGTH: 21
; TYPE: PRT

; ORGANISM: Hepatitis C virus
US-09-995-808-74

Query Match 100.0%; Score 21; DB 25; Length 21;
Best Local Similarity 100.0%; Pred. No. 6.8e-15; Mismatches 0; Indels 0; Gaps 0;
Matches 21; Conservative 0;

Qy 1 LNCNESLNTGWWLAGLIYQHK 21
| | | | | | | | | | | | | | | | | | | | |
Db 1 LNCNESLNTGWWLAGLIYQHK 21

RESULT 7
US-09-995-860-74
; Sequence 74, Application US/09995860
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 74
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-74

Query Match 100.0%; Score 21; DB 25; Length 21;
Best Local Similarity 100.0%; Pred. No. 6.8e-15; Mismatches 0; Indels 0; Gaps 0;
Matches 21; Conservative 0;

Qy 1 LNCNESLNTGWWLAGLIYQHK 21
| | | | | | | | | | | | | | | | | | | | |
Db 1 LNCNESLNTGWWLAGLIYQHK 21

RESULT 8
US-10-020-510-74
; Sequence 74, Application US/10020510
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-72
; CURRENT APPLICATION NUMBER: US/10/020,510
; CURRENT FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 74
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-020-510-74

Query Match 100.0%; Score 21; DB 25; Length 21;
Best Local Similarity 100.0%; Pred. No. 6.8e-15; Mismatches 0; Indels 0; Gaps 0;
Matches 21; Conservative 0;

Qy 1 LNCNESLNTGWWLAGLIYQHK 21
| | | | | | | | | | | | | | | | | | | | |
Db 1 LNCNESLNTGWWLAGLIYQHK 21

RESULT 9
US-10-321-798-74
; Sequence 74, Application US/10321798
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/10/321,798
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 74
; LENGTH: 21
; TYPE: PRT

FILE REFERENCE: 2551-93
CURRENT APPLICATION NUMBER: US/10/321,798
CURRENT FILING DATE: 2002-12-18
PRIOR APPLICATION NUMBER: 60/418,358
PRIOR FILING DATE: 2002-10-16
PRIOR APPLICATION NUMBER: 10/020,510
PRIOR FILING DATE: 2001-12-18
NUMBER OF SEQ ID NOS: 123
SOFTWARE: PatentIn 3.1
SEQ ID NO 74
LENGTH: 21
TYPE: PRT
ORGANISM: Hepatitis C virus
US-10-321-798-74

Query Match 100.0%; Score 21; DB 29; Length 21;
Best Local Similarity 100.0%; Pred. No. 6.9e-15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNLSLNTGWLGLIYQHK 21
Db 1 LNCNLSLNTGWLGLIYQHK 21

RESULT 10

US-08-974-685-30
Sequence 30, Application US/08974685
GENERAL INFORMATION:
APPLICANT: LEROUIX-ROELS, GEERT
DELEYS, ROBERT
MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 181

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,685

FILING DATE: 19-Nov-1997

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-11

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 30:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 30:

Query Match 52.4%; Score 11; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.00033;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNLSLNTGW 11

Db 1 LNCNLSLNTGW 11

RESULT 11

US-08-974-690-30
Sequence 30, Application US/08974690
GENERAL INFORMATION:
APPLICANT: LEROUIX-ROELS, GEERT
DELEYS, ROBERT
MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 166

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,690

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/635,886

FILING DATE: 25-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-11

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 30:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-974-690-30

Query Match 52.4%; Score 11; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.00033;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNLSLNTGW 11

Db 1 LNCNLSLNTGW 11

RESULT 12

US-08-974-690A-30
Sequence 30, Application US/08974690A
GENERAL INFORMATION:
APPLICANT: LEROUIX-ROELS, GEERT
DELEYS, ROBERT
MAERTENS, GEERT

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF

HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 177

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690A
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-08-974-690A-30
Query Match 52.4%; Score 11; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.00033;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 LNCNESLNTGW 11
Db 1 LNCNESLNTGW 11
RESULT 13
US-08-974-690B-30
Sequence 30, Application US/08974690B
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
DELEYS, ROBERT
MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690B
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-08-974-690B-30
Query Match 52.4%; Score 11; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.00033;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 LNCNESLNTGW 11
Db 1 LNCNESLNTGW 11
RESULT 14
PCT-US02-17374-24
Sequence 24, Application PC/TUS0217374
GENERAL INFORMATION:
APPLICANT: Acambis, Inc.
APPLICANT: Kleanthous, Harold
APPLICANT: Oros, Larissa
APPLICANT: Miller, Charles
TITLE OF INVENTION: Chimeric Flavivirus Vectors
FILE REFERENCE: 06132/063WO2
CURRENT APPLICATION NUMBER: PCT/US02/17374
CURRENT FILING DATE: 2003-02-26
PRIOR APPLICATION NUMBER: 60/295,265
PRIOR FILING DATE: 2001-06-01
NUMBER OF SEQ ID NOS: 42
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 24
LENGTH: 27
TYPE: PRT
ORGANISM: Hepatitis C Virus
PCT-US02-17374-24
Query Match 52.4%; Score 11; DB 1; Length 27;
Best Local Similarity 100.0%; Pred. No. 0.00042;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 LNCNESLNTGW 11
Db 17 LNCNESLNTGW 27
RESULT 15
US-08-475-482-93
Sequence 93, Application US/08475482
GENERAL INFORMATION:
APPLICANT: Chang Yi Wang and Barbara Hosein
TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
INFECTION AND PREVENTION THEREOF AS VACCINES
NUMBER OF SEQUENCES: 136
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVE.
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,482

FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/262,037
FILING DATE: 17-June-1994
APPLICATION NUMBER: 07/719,819
FILING DATE: 24-June-1991
APPLICATION NUMBER: 07/667,275
FILING DATE: 11-Mar-1991
APPLICATION NUMBER: 07/651,735
FILING DATE: 07-Feb-1991
APPLICATION NUMBER: 07/558,799
FILING DATE: 26-July-1990
APPLICATION NUMBER: 07/510,153
FILING DATE: 16-April-1990
ATTORNEY/AGENT INFORMATION:
NAME: Maria C. H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4043 US10
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 93:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: Amino acid
STRANDEDNESS:
TOPOLOGY: Unknown
US-08-475-482-93

Query Match 52.4%; Score 11; DB 8; Length 38;
Best Local Similarity 100.0%; Pred. No. 0.00057;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 LNCNESLNTGW 11
Db 6 LNCNESLNTGW 16

Search completed: November 21, 2003, 22:09:50
Job time : 172.255 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.9775 Seconds
(without alignments)
117.536 Million cell updates/sec

Title: US-09-973-025-74
Perfect score: 21
Sequence: 1 LNCNESLNTGWLGLIYQHK 21

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0
Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents AA New:
1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	21	6	US-10-651-165-30
2	11	52.4	30	6	US-10-296-734-464
3	11	52.4	450	6	US-10-651-165-179
4	11	52.4	450	6	US-10-651-165-181
5	11	52.4	1997	6	US-10-296-734-816
6	11	52.4	2010	6	US-10-296-734-814
7	11	52.4	3011	6	US-10-296-734-406
8	11	52.4	5985	6	US-10-296-734-810
9	10	47.6	450	6	US-10-651-165-191
10	10	47.6	450	6	US-10-651-165-192
11	10	47.6	2280	1	PCT-US03-20322-211
12	8	38.1	20	6	US-10-651-165-31
13	7	33.3	30	6	US-10-296-734-462
14	7	33.3	181	6	US-10-425-114A-60359
15	6	28.6	276	5	US-10-679-063-10991
16	6	28.6	276	5	US-09-897-516A-6433
17	6	28.6	333	1	PCT-US03-19834-3
18	6	28.6	333	6	US-10-655-562-4
19	6	28.6	347	6	US-10-664-351-9
20	6	28.6	382	5	US-09-581-286A-491
21	6	28.6	447	5	US-09-581-286A-357
22	6	28.6	450	6	US-10-651-165-180
23	6	28.6	450	6	US-10-651-165-194
24	6	28.6	498	6	US-10-679-063-8664
25	6	28.6	539	6	US-10-664-391-11
26	6	28.6	549	6	US-10-679-063-16221

27	6	28.6	637	1	PCT-US03-33610-4	Sequence 4, Appli
28	6	28.6	3011	1	PCT-US03-19834-2	Sequence 2, Appli
29	5	23.8	17	5	US-09-538-038A-1971	Sequence 1971, Ap
30	5	23.8	17	5	US-09-538-038A-1984	Sequence 1984, Ap
31	5	23.8	17	5	US-09-538-038A-1985	Sequence 1985, Ap
32	5	23.8	17	5	US-09-538-038A-1986	Sequence 1986, Ap
33	5	23.8	17	5	US-09-538-038A-1988	Sequence 1988, Ap
34	5	23.8	17	5	US-09-538-038A-1993	Sequence 1993, Ap
35	5	23.8	17	5	US-09-538-038A-1995	Sequence 1995, Ap
36	5	23.8	17	5	US-09-538-038A-2001	Sequence 2001, Ap
37	5	23.8	17	5	US-09-538-038A-2004	Sequence 2004, Ap
38	5	23.8	17	5	US-09-538-038A-2308	Sequence 2308, Ap
39	5	23.8	17	5	US-09-962-756A-1542	Sequence 1542, Ap
40	5	23.8	17	6	US-10-253-471A-1542	Sequence 1542, Ap
41	5	23.8	23	5	US-09-538-038A-2114	Sequence 2114, Ap
42	5	23.8	23	5	US-09-538-038A-2413	Sequence 2413, Ap
43	5	23.8	23	5	US-09-962-756A-1170	Sequence 1170, Ap
44	5	23.8	23	5	US-09-962-756A-1541	Sequence 1541, Ap
45	5	23.8	23	6	US-10-253-471A-1170	Sequence 1170, Ap

ALIGNMENTS

RESULT 1
US-10-651-165-30
; Sequence 30, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 21
; TYPE: PRP
; ORGANISM: hepatitis C virus
US-10-651-165-30

Query Match 100.0%; Score 21; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.7e-14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 LNCNESLNTGWLGLIYQHK 21
| | | | | | | | | | | | | | | | | | | | | |
Db 1 LNCNESLNTGWLGLIYQHK 21
| | | | | | | | | | | | | | | | | | | | | |

RESULT 2
US-10-296-734-464
; Sequence 464, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2

```
; SEQ ID NO 464
; LENGTH: 30
; TYPE: PRT
; ORGANISM: hepatitis C virus
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 29
US-10-296-734-464

Query Match      52.4%; Score 11; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 LNCNESLNTGW 11
      |||||
Db      9 LNCNESLNTGW 19

RESULT 3
US-10-651-165-179
; Sequence 179, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 179
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-179

Query Match      52.4%; Score 11; DB 6; Length 450;
Best Local Similarity 100.0%; Pred. No. 0.00076;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 LNCNESLNTGW 11
      |||||
Db      427 LNCNESLNTGW 437

RESULT 4
US-10-651-165-181
; Sequence 181, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 181
```

```
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-181

Query Match      52.4%; Score 11; DB 6; Length 450;
Best Local Similarity 100.0%; Pred. No. 0.00076;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 LNCNESLNTGW 11
      |||||
Db      427 LNCNESLNTGW 437

RESULT 5
US-10-296-734-816
; Sequence 816, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 816
; LENGTH: 1997
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette C
US-10-296-734-816

Query Match      52.4%; Score 11; DB 6; Length 1997;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 LNCNESLNTGW 11
      |||||
Db      594 LNCNESLNTGW 604

RESULT 6
US-10-296-734-814
; Sequence 814, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 814
; LENGTH: 2010
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette B
US-10-296-734-814

Query Match      52.4%; Score 11; DB 6; Length 2010;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 LNCNESLNTGW 11
      |||||
```

Db 1359 LNCNLSLNTGW 1369

RESULT 7

US-10-296-734-406
; Sequence 406, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la consensus polyprotein
US-10-296-734-406

Query Match 52.4%; Score 11; DB 6; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.0026;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNLSLNTGW 11
Db 427 LNCNLSLNTGW 437

RESULT 8

US-10-296-734-810
; Sequence 810, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 810
; LENGTH: 5985
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la savine
US-10-296-734-810

Query Match 52.4%; Score 11; DB 6; Length 5985;
Best Local Similarity 100.0%; Pred. No. 0.0042;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNLSLNTGW 11
Db 3369 LNCNLSLNTGW 3379

RESULT 9

US-10-651-165-191
; Sequence 191, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C

; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 191
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-191

Query Match 47.6%; Score 10; DB 6; Length 450;
Best Local Similarity 100.0%; Pred. No. 0.0072;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNLSLNTG 10
Db 427 LNCNLSLNTG 436

RESULT 10

US-10-651-165-192
; Sequence 192, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 192
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-192

Query Match 47.6%; Score 10; DB 6; Length 450;
Best Local Similarity 100.0%; Pred. No. 0.0072;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNLSLNTG 10
Db 427 LNCNLSLNTG 436

RESULT 11

PCT-US03-20322-211
; Sequence 211, Application PC/TUS0320322
; GENERAL INFORMATION:
; APPLICANT: Dana-Farber Cancer Institute, Inc., et al.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
; FILE REFERENCE: DFN-043PC
; CURRENT APPLICATION NUMBER: PCT/US03/20322
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718

```
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 211
; LENGTH: 2280
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
PCT-US03-20322-211

Query Match      47.6%; Score 10; DB 1; Length 2280;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 LNCNESLNTG 10
Db      427 LNCNESLNTG 436

RESULT 12
US-10-651-165-31
; Sequence 31, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651.165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974.690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-31

Query Match      38.1%; Score 8; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.082;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      14 AGLIYQHK 21
Db      1 AGLIYQHK 8

RESULT 13
US-10-296-734-462
; Sequence 462, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296.734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 462
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 28

US-10-296-734-462
Query Match      33.3%; Score 7; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 LNCNESL 7
Db      24 LNCNESL 30

RESULT 14
US-10-425-114A-60359
; Sequence 60359, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425.114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 60359
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3607-047-A6_FLI.pcp
US-10-425-114A-60359

Query Match      33.3%; Score 7; DB 6; Length 181;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 LNCNESL 7
Db      27 LNCNESL 33

RESULT 15
US-10-679-063-10991
; Sequence 10991, Application US/10679063
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52054)B
; CURRENT APPLICATION NUMBER: US/10/679.063
; CURRENT FILING DATE: 2003-10-02
; PRIOR APPLICATION NUMBER: 60/415,758
; PRIOR FILING DATE: 2002-10-02
; NUMBER OF SEQ ID NOS: 27373
; SEQ ID NO 10991
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-10-679-063-10991

Query Match      28.6%; Score 6; DB 6; Length 274;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 ESLNTG 10
Db      257 ESLNTG 262

Search completed: November 21, 2003, 22:12:53
Job time : 8.9775 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:46:27 ; Search time 9.7125 Seconds
(without alignments)
207.932 Million cell updates/sec

Title: US-09-973-025-74

Perfect score: 21
Sequence: 1 LNCNESLNTGWLAGLIYQHK 21

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 76.*

- 1: pir1.*
- 2: pir2.*
- 3: pir3.*
- 4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	52.4	138	2 S24102	envelope protein -
2	11	52.4	138	2 S24069	envelope protein -
3	11	52.4	3011	1 GNWVCH	genome polyprotein
4	11	52.4	3011	1 S40770	genome polyprotein
5	10	47.6	134	2 S24089	envelope protein -
6	10	47.6	3010	1 A45573	genome polyprotein
7	7	33.3	138	2 S24085	envelope protein -
8	7	33.3	138	2 S24101	envelope protein -
9	7	33.3	138	2 S24068	envelope protein -
10	7	33.3	315	2 P80164	envelope glycoprot
11	7	33.3	716	2 JQ1366	polyprotein - hepa
12	6	28.6	62	2 S28524	tonoplast intrinsi
13	6	28.6	62	2 T13997	aquaporin - common
14	6	28.6	91	2 C51627	fibroblast growth
15	6	28.6	114	2 D84766	membrane protein y
16	6	28.6	114	2 F90682	hypothetical prote
17	6	28.6	114	2 B85533	hypothetical prote
18	6	28.6	138	2 S24080	envelope protein -
19	6	28.6	138	2 S24081	envelope protein -
20	6	28.6	175	2 B70783	hypothetical prote
21	6	28.6	190	2 T11321	NADH2 dehydrogenas
22	6	28.6	217	2 I51062	MHC class II beta
23	6	28.6	239	2 B64509	hypothetical prote
24	6	28.6	238	2 T21050	hypothetical prote
25	6	28.6	244	2 S29982	class II histocomp
26	6	28.6	245	2 S29980	class II histocomp
27	6	28.6	247	2 I51059	MHC class II beta
28	6	28.6	247	2 I51060	MHC class II beta
29	6	28.6	248	2 T14002	aquaporin TIP7 - c

ALIGNMENTS

RESULT 1

S24102

envelope protein - hepatitis C virus (fragment)

N:Contains: envelope protein gp35; envelope protein gp70

C:Species: hepatitis C virus

C>Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000

C:Accession: S24102

R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraishi, K.; Ohkoshi,

Virus Res. 22, 107-123, 1992

A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus

A:Reference number: S24068; MUID:92230394; PMID:1314471

A:Accession: S24102

A:Molecule type: genomic RNA

A:Residues: 1-138 <KAT>

A:Cross-references: EMBL:X60585

A:Experimental source: isolate REB

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: envelope protein; glycoprotein; polyprotein

F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>

F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>

F:16-42/Region: hypervariable 1 #status predicted

F:106-112/Region: hypervariable 2 #status predicted

F:49,55,62,80,110/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 52.4%; Score 11; DB 2; Length 138;

Best Local Similarity 100.0%; Pred. No. 8.5e-05;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11

Db 59 LNCNESLNTGW 69

RESULT 2

S24069

envelope protein - hepatitis C virus (fragment)

N:Contains: envelope protein gp35; envelope protein gp70

C:Species: hepatitis C virus

C>Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000

C:Accession: S24069

R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraishi, K.; Ohkoshi,

Virus Res. 22, 107-123, 1992

A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus

A:Reference number: S24068; MUID:92230394; PMID:1314471

A:Accession: S24069

A:Molecule type: genomic RNA

A:Residues: 1-138 <KAT>

A:Cross-references: EMBL:X60553

A:Experimental source: isolate RE1B

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: envelope protein; glycoprotein; polyprotein

F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>

F;16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
 F;16-42/Region: hypervariable 1 #status predicted
 F;106-112/Region: hypervariable 2 #status predicted
 F;49,55,62,80,110/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 52.4%; Score 11; DB 2; Length 138;
 Best Local Similarity 100.0%; Pred. No. 8.5e-05;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
 |||||
 Db 59 LNCNESLNTGW 69

RESULT 3
 GNWVCH
 Genome polyprotein - hepatitis C virus (strain H)
 N;Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C;Species: hepatitis C virus
 A;Note: host Homo sapiens (man)
 C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C;Accession: A36814; A41546
 R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 A;Description: Genomic structure of the human prototype strain H of hepatitis C virus: compared to GenBank, July 1992
 A;Reference number: A36814
 A;Accession: A36814
 A;Molecule type: genomic RNA
 A;Residues: 1-3011 <INC>
 A;Cross-references: GB:M67463; NID:G329737; PIDN:AAA45534.1; PID:G329738
 R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
 A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: compared to GenBank, July 1992
 A;Reference number: A41546; MUID:92052256; PMID:1659800
 A;Contents: annotation
 A;Note: neither amino acid nor nucleotide sequence is given
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein
 F;1-15/Product: capsid protein C #status predicted <CPC>
 F;116-191/Product: envelope protein M #status predicted <EPM>
 F;192-389/Product: major envelope protein E #status predicted <MEE>
 F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F;1007-1615/Product: hepatitis C virus genome polyprotein
 F;1230-1237/Region: nucleotide-binding motif A (P-loop)
 F;1312-1317/Region: nucleotide-binding motif B
 F;1316-1319/Region: DEXH motif
 F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F;1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>
 F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match 52.4%; Score 11; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
 |||||
 Db 427 LNCNESLNTGW 437

RESULT 4
 S40770
 Genome polyprotein - hepatitis C virus
 N;Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C;Species: hepatitis C virus
 C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C;Accession: S40770; FCI285
 R;Okamoto, H.
 submitted to the EMBL Data Library, March 1992
 A;Reference number: S40770

A;Accession: S40770
 A;Molecule type: genomic RNA
 A;Residues: 1-3011 <OKA>
 A;Cross-references: EMBL:D10749; NID:G221586; PIDN:BAA01582.1; PID:G221587
 R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990
 A;Title: The 5'-terminal sequence of the hepatitis C virus genome.
 A;Reference number: FCI284; MUID:91013116; PMID:2170712
 A;Accession: FCI285
 A;Molecule type: genomic RNA
 A;Residues: 1-513 <OK2>
 A;Cross-references: GB:D00831; NID:G221511; PIDN:BAA00705.1; PID:G221512
 A;Experimental source: isolate HC-J1
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serine
 F;2-115/Product: capsid protein C #status predicted <CPC>
 F;116-191/Product: envelope protein M #status predicted <EPM>
 F;192-389/Product: major envelope protein E #status predicted <MEE>
 F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F;1007-1615/Product: hepatitis C virus genome polyprotein
 F;1230-1237/Region: nucleotide-binding motif A (P-loop)
 F;1312-1317/Region: nucleotide-binding motif B
 F;1316-1319/Region: DEXH motif
 F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 52.4%; Score 11; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
 |||||
 Db 427 LNCNESLNTGW 437

RESULT 5
 S24089
 envelope protein - hepatitis C virus (fragment)
 N;Contains: envelope protein gp35; envelope protein gp70
 C;Species: hepatitis C virus
 C;Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
 C;Accession: S24089; S78230
 R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraishi, K.; Ohkoshi, Virus Res. 22, 107-123, 1992
 A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
 A;Reference number: S24068; MUID:92230394; PMID:1314471
 A;Accession: S24089
 A;Molecule type: genomic RNA
 A;Residues: 1-134 <KAT>
 A;Cross-references: EMBL:X60575
 A;Experimental source: isolate RES5D
 R;Kato, N.
 submitted to the EMBL Data Library, August 1991
 A;Reference number: S78226
 A;Accession: S78230
 A;Molecule type: genomic RNA
 A;Residues: 1-15, 'XXXX', 16-134 <KAW>
 A;Cross-references: EMBL:X60575
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: envelope protein; glycoprotein; polypeptide
 F;1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
 F;16-134/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
 F;16-38/Region: hypervariable 1 #status predicted
 F;102-108/Region: hypervariable 2 #status predicted
 F;45,51,58,76/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 47.6%; Score 10; DB 2; Length 134;
 Best Local Similarity 100.0%; Pred. No. 0.00097;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 10

Db 55 LNCNESLNTG 64
|||||||

RESULT 6

A45573
Genome polyprotein - hepatitis C virus (strain J7)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (fragment)
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: A45573
R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, Y.
Virus Res. 23, 39-53, 1992
A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: a
A:Reference number: A45573; MUID:92295714; PMID:1318627
A:Accession: A45573
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3010 <TAN>
A:Cross-references: GB:D11168; GB:D01171; NID:g221612; PIDN:BAA01943.1; PID:g221613
A:Experimental source: HCV-JT
A:Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBI:P106207)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <BPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DRX motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 47.6%; Score 10; DB 1; Length 3010;

Best Local Similarity 100.0%; Pred. No. 0.015;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTG 10

|||||||

Db 427 LNCNESLNTG 436

RESULT 7

S24085
envelope protein - hepatitis C virus (fragment)
N:Contains: envelope protein gp35; envelope protein gp70
C:Species: hepatitis C virus
C:Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C:Accession: S24085

R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraisho, K.; Ohkoshi, Y.

Virus Res. 22, 107-123, 1992
A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus

A:Reference number: S24068; MUID:92230394; PMID:1314471

A:Accession: S24085

A:Molecule type: genomic RNA

A:Residues: 1-138 <KAT>

A:Cross-references: EMBL:X60571

A:Experimental source: isolate RE54

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: envelope protein; glycoprotein; polyprotein

F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>

F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>

F:16-42/Region: hypervariable 1 #status predicted

F:106-112/Region: hypervariable 2 #status predicted

F:49,55,62,80,110/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match

Best Local Similarity 33.3%; Score 7; DB 2; Length 138;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESL 7

|||||||

Db 59 LNCNESL 65

RESULT 8

S24101
envelope protein - hepatitis C virus (fragment)
N:Contains: envelope protein gp35; envelope protein gp70
C:Species: hepatitis C virus
C:Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C:Accession: S24101
R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraisho, K.; Ohkoshi, Y.

Virus Res. 22, 107-123, 1992

A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus

A:Reference number: S24068; MUID:92230394; PMID:1314471

A:Accession: S24101

A:Molecule type: genomic RNA

A:Residues: 1-138 <KAT>

A:Cross-references: EMBL:X60584

A:Experimental source: isolate RE72A

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: envelope protein; glycoprotein; polyprotein

F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>

F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>

F:16-42/Region: hypervariable 1 #status predicted

F:106-112/Region: hypervariable 2 #status predicted

F:49,55,62,80,110/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match

Best Local Similarity 33.3%; Score 7; DB 2; Length 138;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 LNCNESL 7

|||||||

Db 59 LNCNESL 65

RESULT 9

S24068
envelope protein - hepatitis C virus (fragment)
N:Contains: envelope protein gp35; envelope protein gp70
C:Species: hepatitis C virus
C:Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C:Accession: S24068

R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraisho, K.; Ohkoshi, Y.

Virus Res. 22, 107-123, 1992
A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus

A:Reference number: S24068; MUID:92230394; PMID:1314471

A:Accession: S24068

A:Molecule type: genomic RNA

A:Residues: 1-138 <KAT>

A:Cross-references: EMBL:X60552

A:Experimental source: isolate RE1A

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: envelope protein; glycoprotein; polyprotein

F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>

F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>

F:16-42/Region: hypervariable 1 #status predicted

F:106-112/Region: hypervariable 2 #status predicted

F:49,55,62,80,110/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match

Best Local Similarity 33.3%; Score 7; DB 2; Length 138;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESL 7

|||||||

Db 59 LNCNESL 65

```

RESULT 10
PS0164
envelope glycoprotein (clone 64) - hepatitis C virus (fragment)
C:Species: hepatitis C virus
C>Date: 23-Nov-1991 #sequence_revision 23-Nov-1991 #text_change 17-Nov-2000
C:Accession: PS0164
R:Hijikata, M.; Kato, N.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Shimotohno, K.
Biochem. Biophys. Res. Commun. 175, 220-228, 1991
A:Title: Hypervariable regions in the putative glycoprotein of hepatitis C virus.
A:Reference number: FN0011; MUID:91151353; PMID:1847805
A:Accession: PS0164
A:Molecule type: genomic RNA
A:Residues: 1-315 <HI>
A:Cross-references: GB:D00690
A:Note: the authors translated the codon GAC for residues 27 and 112 as Glu
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: glycoprotein; polyprotein
F:200-209/Region: hypervariable 1 #status predicted
F:283-289/Region: hypervariable 2 #status predicted
F:5,18,43,59,114,134,226,232,239,257,287/Binding site: carbohydrate (Asn) (covalent) #st
P:5,18,43,59,114,134,226,232,239,257,287/Binding site: carbohydrate (Asn) (covalent) #st

Query Match 33.3%; Score 7; DB 2; Length 315;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESL 7
Db 236 LNCNESL 242

RESULT 11
JQ1366
polyprotein - hepatitis C virus (French isolate) (fragments)
C:Species: hepatitis C virus
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
R:Kremsdorff, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.
J. Gen. Virol. 72, 2557-2561, 1991
A:Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication
A:Reference number: JQ1366; MUID:92013977; PMID:1655961
A:Accession: JQ1366
A:Molecule type: genomic RNA
A:Residues: 1-716 <KR>
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: glycoprotein; polyprotein
F:84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #stac

Query Match 33.3%; Score 7; DB 2; Length 716;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESL 7
Db 94 LNCNESL 100

RESULT 12
S28524
tonoplast intrinsic protein isoform gamma - Arabidopsis thaliana (fragment)
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 21-Aug-1998
C:Accession: S28524
R:Desprez, T.; Amealem, J.; Chippello, H.; Rouze, P.; Caboche, M.; Hofte, H.
submitted to the EMBL Data Library, November 1992
A:Description: The Arabidopsis thaliana transcribed genome: the GDR cDNA program.
A:Reference number: S28524
A:Accession: S28524
A:Molecule type: mRNA
A:Residues: 1-62 <DES>
A:Cross-references: EMBL:Z18127
C:Superfamily: lens fiber membrane major intrinsic protein

Query Match 28.6%; Score 6; DB 2; Length 62;

```

```

Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 LAGLIY 18
Db 39 LAGLIY 44

RESULT 13
aquaporin - common sunflower (fragment)
N:Alternate names: TIP-like protein; tonoplast intrinsic protein homolog
C:Species: Helianthus annuus (common sunflower)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Jun-2000
C:Accession: T13997
R:Sarda, X.; Tusch, D.; Ferrare, K.; Legrand, E.; Dupuis, J.M.; Casse-Delbart, F.; Lama
Plant J. 12, 1103-1111, 1997
A:Title: Two TIP-like genes encoding aquaporins are expressed in sunflower guard cells.
A:Reference number: Z17561; MUID:98079246; PMID:9418051
A:Accession: T13997
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-62 <SAR>
A:Cross-references: EMBL:X95954; NID:G1212918
C:Superfamily: lens fiber membrane major intrinsic protein
C:Keywords: channel-forming protein; transmembrane protein

Query Match 28.6%; Score 6; DB 2; Length 62;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 LAGLIY 18
Db 41 LAGLIY 46

RESULT 14
C53627
fibroblast growth factor receptor 3, soluble form - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 30-May-1997
C:Accession: C53627
R:Chellaiiah, A.T.; McEwen, D.G.; Werner, S.; Xu, J.; Ornitz, D.M.
J. Biol. Chem. 269, 11620-11627, 1994
A:Title: Fibroblast growth factor receptor (FGFR) 3. Alternative splicing in immunoglobu
A:Reference number: A53627; MUID:94209351; PMID:7512569
A:Accession: oC53627
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-91 <CHE>
A:Cross-references: GB:L26492
C:Keywords: growth factor receptor

Query Match 28.6%; Score 6; DB 2; Length 91;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 GWMLAG 15
Db 69 GWMLAG 74

RESULT 15
D64766
membrane protein yaiz - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: D64766
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503

```

A;Accession: D64766
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-114 <BLAT>
A;Cross-references: GB:AE000144; GB:U00096; NID:g1786568; PIDN:AAC73483.1; PID:g1786578;
A;Experimental source: strain K-12, substrain MG1655

C;Genetics:
A;Gene: yaiz
C;Superfamily: Escherichia coli membrane protein yaiz
C;Keywords: transmembrane protein
F;59-75/Domain: transmembrane #status predicted <TM1>
F;87-103/Domain: transmembrane #status predicted <TM2>

Query Match 28.6%; Score 6; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 WLAGLI 17
Db 97 WLAGLI 102

Search completed: November 21, 2003, 21:11:29
Job time : 9.7125 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:52 ; Search time 5.0925 Seconds
(without alignments)
193.925 Million cell updates/sec

Title: US-09-973-025-74

Perfect score: 21

Sequence: 1 LNCNLSLNTGWLGLYQHK 21

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	52.4	3011	1 POLG_HCVH	P27958 h genome po
2	10	47.6	3010	1 POLG_HCVH	Q00269 h genome po
3	6	28.6	70	1 YAI2_ECOLI	P77273 escherichia
4	6	28.6	138	1 POLG_HCVH	P27953 hepatitis c
5	6	28.6	146	1 NUOA_ERWCA	O85273 erwinia car
6	6	28.6	175	1 Y501_MYCTU	Q10559 mycobacteri
7	6	28.6	235	1 SCE2_YEAST	P03878 saccharomyc
8	6	28.6	238	1 YG76_METUA	Q59070 methanococc
9	6	28.6	260	1 RCEH_RHOSH	P11846 rhodobacter
10	6	28.6	274	1 UPK1_BACHD	Q9kfl5 bacillus ha
11	6	28.6	278	1 YLMD_BACSU	O31726 bacillus su
12	6	28.6	316	1 THER_BACTH	P00800 bacillus th
13	6	28.6	321	1 POLG_HCVH	P27956 hepatitis c
14	6	28.6	321	1 POLG_HCVH	P27957 hepatitis c
15	6	28.6	334	1 MTN3_NEILA	P24582 neisseria l
16	6	28.6	384	1 YXT2_YEAST	P03879 saccharomyc
17	6	28.6	390	1 UL33_HCMVA	P16849 human cytom
18	6	28.6	551	1 NPRS_BACST	P43133 bacillus st
19	6	28.6	3011	1 POLG_HCVI	P26664 h genome po
20	5	23.8	37	1 POLN_WEEV	P13896 western equ
21	5	23.8	48	1 FMBE_BACNO	P17828 bacteroides
22	5	23.8	48	1 FMBE_BACNO	P17831 bacteroides
23	5	23.8	48	1 FMBF_BACNO	P17832 bacteroides
24	5	23.8	48	1 FMBG_BACNO	P17833 bacteroides
25	5	23.8	48	1 TXA2_RADMA	P30783 radianthus
26	5	23.8	51	1 YF93_HAEIN	P44264 haemophilus
27	5	23.8	52	1 CMDD_RHOCA	P29963 rhodobacter
28	5	23.8	69	1 Y842_ARCFU	O29416 archaeoglob
29	5	23.8	82	1 COAB_BPFP1	P03621 bacterioph
30	5	23.8	90	1 REV_F1W28	P05869 human immun
31	5	23.8	92	1 GONI_HUMAN	P01148 homo sapien
32	5	23.8	92	1 GONI_TUPGB	Q95335 tupai
33	5	23.8	93	1 CD3G_BOVIN	Q28074 bos taurus

RESULT 1

ID	POLG_HCVH	STANDARD	PRT	3011 AA
AC	P27958			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].			
DE	Hepatitis C virus (isolate H) (HCV)			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.			
OX	NCBI_TaxID=11108;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92052256; PubMed=1658800;			
RA	Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M., Prince A.M.;			
RT	"Genomic structure of the human prototype strain H of hepatitis C virus: comparison with American and Japanese isolates.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).			
RN	[2]			
RP	X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.			
RX	MEDLINE=97331322; PubMed=9187654;			
RA	Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;			
RT	"Structure of the hepatitis C virus RNA helicase domain.";			
RL	Nat. Struct. Biol. 4:463-467(1997).			
RN	[3]			
RP	X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.			
RX	MEDLINE=98154321; PubMed=9493270;			
RA	Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A., Murcko M.A., Lin C., Caron P.R.;			
RT	"Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding.";			
RL	Structure 6:89-100(1998).			
CC	-1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.			
CC	-1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.			
CC	-1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE ACTIVATION OF NS3.			
CC	-1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.			
CC	-1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN ESSENTIAL ROLE IN THE VIRUS REPLICATION.			
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	-1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).			
CC	-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1			

CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -1- PTM: THE STRUCTURAL PROTEINS C, B1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M67463; AAA5534.1; --
 CC PIR; A36814; GNVVCH.
 CC PDB; 1HEI; 25-NOV-98.
 CC PDB; 1A1V; 16-FEB-99.
 CC PDB; 1A1R; 17-JUN-98.
 CC MEROPS; S29.001; --
 CC MEROPS; U39.001; --
 CC TRANSFAC; T04155; -- DEAD.
 CC InterPro; IPR001410; HCV capsid.
 CC InterPro; IPR002522; HCV capsid.
 CC InterPro; IPR002521; HCV core.
 CC InterPro; IPR002519; HCV env.
 CC InterPro; IPR002531; HCV NS1.
 CC InterPro; IPR002518; HCV NS2.
 CC InterPro; IPR004109; HCV NS3.
 CC InterPro; IPR000745; HCV NS4a.
 CC InterPro; IPR001490; HCV NS4b.
 CC InterPro; IPR002868; HCV NS5a.
 CC InterPro; IPR002166; HCV RdRp.
 CC InterPro; IPR001650; Helicase C.
 CC InterPro; IPR007095; RNA pol DS_PS.
 CC InterPro; IPR007094; RNA pol Psvir.
 CC Pfam; PF01543; HCV capsid; 1.
 CC Pfam; PF01542; HCV core; 1.
 CC Pfam; PF01539; HCV env; 1.
 CC Pfam; PF01560; HCV NS1; 1.
 CC Pfam; PF01538; HCV NS2; 1.
 CC Pfam; PF02907; HCV NS3; 1.
 CC Pfam; PF01006; HCV NS4a; 1.
 CC Pfam; PF01001; HCV NS4b; 1.
 CC Pfam; PF01506; HCV NS5a; 1.
 CC Pfam; PF00271; Helicase C; 1.
 CC Pfam; PF00998; Viral RdRp; 1.
 CC ProDom; PD186062; HCV NS1; 1.
 CC SMART; SM00487; DEXDc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 192 383 CAPSID PROTEIN C.
 FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
 FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
 FT CHAIN 810 1026 PROTEIN P7.
 FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
 FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.
 FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4A.
 FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4B.
 FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5A.
 FT CHAIN 3011 369 NONSTRUCTURAL PROTEIN NS5B.
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT 234 234 CARBOHYD
 FT 305 305 CARBOHYD
 FT 417 417 CARBOHYD
 FT 423 423 CARBOHYD
 FT 430 430 CARBOHYD
 FT 448 448 CARBOHYD
 FT 476 476 CARBOHYD
 FT 532 532 CARBOHYD
 FT 540 540 CARBOHYD
 FT 556 556 CARBOHYD
 FT 576 576 CARBOHYD
 FT 623 623 CARBOHYD
 FT 645 645 CARBOHYD
 FT STRAND 1224 1226
 FT TURN 1232 1233
 FT TURN 1236 1238
 FT TURN 1239 1246
 FT TURN 1247 1248
 FT STRAND 1251 1255
 FT HELIX 1258 1271
 FT TURN 1272 1272
 FT STRAND 1277 1280
 FT TURN 1281 1282
 FT STRAND 1283 1285
 FT STRAND 1291 1295
 FT HELIX 1296 1301
 FT TURN 1302 1303
 FT STRAND 1312 1316
 FT TURN 1317 1319
 FT HELIX 1323 1335
 FT TURN 1336 1340
 FT STRAND 1343 1347
 FT TURN 1352 1353
 FT TURN 1360 1361
 FT STRAND 1362 1366
 FT STRAND 1368 1368
 FT STRAND 1373 1375
 FT TURN 1376 1377
 FT STRAND 1378 1380
 FT HELIX 1382 1385
 FT STRAND 1389 1393
 FT HELIX 1397 1409
 FT TURN 1410 1411
 FT STRAND 1414 1417
 FT TURN 1419 1420
 FT STRAND 1432 1436
 FT TURN 1438 1439
 FT STRAND 1450 1453
 FT STRAND 1456 1463
 FT STRAND 1471 1478
 FT STRAND 1480 1480
 FT HELIX 1481 1488
 FT TURN 1489 1490
 FT STRAND 1497 1501
 FT STRAND 1507 1507
 FT STRAND 1511 1511
 FT HELIX 1514 1527
 FT STRAND 1532 1544
 FT STRAND 1550 1550
 FT HELIX 1555 1564
 FT HELIX 1570 1578
 FT TURN 1579 1580
 FT HELIX 1584 1597
 FT TURN 1598 1598
 FT HELIX 1606 1611
 FT TURN 1614 1618
 FT STRAND 1622 1623
 FT STRAND 1627 1627
 FT STRAND 1635 1636
 FT HELIX 1640 1652
 FT SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;
 Query Match 52.4%; Score 11; DB 1; Length 3011;

Best Local Similarity 100.0%; Pred. No. 0.00072;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTG 11
Db 427 LNCNESLNTG 437

RESULT 2
POLG_HCVJT STANDARD; PRT; 3010 AA.

AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals."
RL Virus Res. 23:39-53(1992).

CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC [RNA] (N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MENA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D11168; BAA01943.1; -
CC PIR; A45573; A45573.
CC PDB; 1AIQ; 25-MAR-98.
CC PDB; LUXP; 14-JAN-98.
CC MEROPS; S29.001; -
CC MEROPS; U39.001; -
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002522; HCV_capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC InterPro; IPR002531; HCV_NS1.
CC InterPro; IPR002518; HCV_NS2.
CC InterPro; IPR004109; HCV_NS3.
CC InterPro; IPR000745; HCV_NS4a.
CC InterPro; IPR001490; HCV_NS4b.
CC InterPro; IPR002868; HCV_NS5a.

DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXdc; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 1007 1615 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT ACT_SITE 1083 1083 POTENTIAL.
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;

Query Match 47.6%; Score 10; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 0.0083;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTG 10
Db 427 LNCNESLNTG 436

RESULT 3
YAIZ_ECOLI
ID YAIZ_ECOLI STANDARD; PRT; 70 AA.
AC P77273;
DT 15-JUL-1998 (Rel. 36, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein yaiz.
 GN YAIZ OR B0380 OR C0486.
 OS Escherichia coli, and
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562, 217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
 RA Federpriel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
 RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=2238234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner P.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 DR EMBL; AE000144; AAC73483.1; ALT_INIT.
 DR EMBL; U73857; AAB18103.1; ALT_INIT.
 DR EMBL; AE016756; AAN78964.1; ALT_INIT.
 DR EcoGene; EG14280; yaiz.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 13 33 POTENTIAL.
 FT TRANSMEM 39 59 POTENTIAL.
 SQ SEQUENCE 70 AA; 8097 MW; 10EB267CCED4BCEA CRC64;
 Query Match 28.6%; Score 6; DB 1; Length 70;
 Best Local Similarity 100.0%; Pred. No. 5.5;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 12 WLGLI 17
 Db 53 WLGLI 58
 RESULT 4
 ID POLG HCVE0 STANDARD; PRT; 138 AA.
 AC P27953;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Envelope glycoprotein E1 (GP32) (GP35);
 DE Envelope glycoprotein E2 (GP68) (GP70) (NS1)] (Fragment).

OS Hepatitis C virus (isolate EC10) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11106;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91112009; PubMed=1846505;
 RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
 RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
 RA Han J.H.;
 RT "Variable and hypervariable domains are found in the regions of HCV
 corresponding to the flavivirus envelope and NS1 proteins and the
 pestivirus envelope glycoproteins";
 RL Virology 180:843-848(1991).
 CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 DR EMBL; X53136; CAA37296.1; -.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT NON TER 1 1
 FT CHAIN <1 84 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 85 >138 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CARBOHYD 6 6 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 118 118 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 124 124 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 131 131 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT NON TER 138 138
 SQ SEQUENCE 138 AA; 14781 MW; CD3FOA962DEAB1AD CRC64;
 Query Match 28.6%; Score 6; DB 1; Length 138;
 Best Local Similarity 100.0%; Pred. No. 9.9;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 SLNTGW 11
 Db 133 SLNTGW 138
 RESULT 5
 ID NUOA ERWCA STANDARD; PRT; 146 AA.
 AC O85273;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE NADH-quinone oxidoreductase chain A (EC 1.6.99.5) (NADH dehydrogenase
 I, chain A) (NDH-1, chain A).
 GN NUOA.
 OS Erwinia carotovora.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pectobacterium.
 OX NCBI_TaxID=554;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ECC193;
 RX MEDLINE=98305692; PubMed=9643539;
 RA Harris S.J., Shih Y.L., Bentley S.D., Salmond G.P.C.;

RT "The hexA gene of *Erwinia carotovora* encodes a lysozyme homologue and
RT regulates motility and the expression of multiple virulence
RT determinants.";
RL Mol. Microbiol. 28:705-717(1998).
CC -|- FUNCTION: NDH-1 shuttles electrons from NADH, via FMN and iron-
CC sulfur (Fe-S) centers, to quinones in the respiratory chain.
CC Couples the redox reaction to proton translocation (for every two
CC electrons transferred, four hydrogen ions are translocated across
CC the cytoplasmic membrane), and thus conserves the redox energy in
CC a proton gradient (by similarity).
CC -|- CATALYTIC ACTIVITY: NADH + quinone = NAD(+) + quinol.
CC -|- SUBUNIT: COMPOSED OF 13 DIFFERENT SUBUNITS. SUBUNIT NUOA, H, J,
CC K, L, M, N CONSTITUTE THE MEMBRANE SECTOR OF THE COMPLEX.
CC -|- SUBCELLULAR LOCATION: Integral membrane protein.
CC -|- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 3 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF057063; AAC38640.1; -.
DR InterPro; IPR000440; Oxidored_q4.
DR Pfam; PF00507; oxidored_q4; 1.
KW Oxidoreductase; NAD; Quinone; Transmembrane.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 68 88 POTENTIAL.
FT TRANSMEM 96 116 POTENTIAL.
SQ SEQUENCE 146 AA; 16156 MW; D6B016B3D9292AA6 CRC64;

Query Match 28.6%; Score 6; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 LAGLIY 18
DB 109 LAGLIY 114

RESULT 6
Y901 MYCTU STANDARD; PRT; 175 AA.
AC Q10559;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein RV0901.
GN RV0901 OR MT0924 OR MTCV31.29.
OS Mycobacterium tuberculosis
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Barry C.E. III, Tekai F.,
RA Gordon S.V., Eigmeier K., Gas S., Chillingworth T., Connor R.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of *Mycobacterium tuberculosis* from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of *Mycobacterium tuberculosis* clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z73101; CA97376.1; -; ALT_INIT.
DR EMBL; AE006979; AAK45171.1; -; ALT_INIT.
DR PIR; B70783; B70783.
DR TIGR; MT0924; -.
DR TuberculList; Rv0901; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 175 AA; 18909 MW; AD7024B10E5E1704 CRC64;

Query Match 28.6%; Score 6; DB 1; Length 175;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 WLAGL 16
DB 6 WLAGL 11

RESULT 7
SCE2_YEAST STANDARD; PRT; 235 AA.
AC P03878;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Intron-encoded endonuclease I-SceII (EC 3.1.1.-) (DNA endonuclease I-
DE SceII) (DNA endonuclease AI4) (PAL4 nuclease).
GN AI4 OR ENS2 OR I-SCEII.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Mitochondrion.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D273-10B;
RX MEDLINE=81069885; PubMed=6254986;
RA Bonitz S.G., Coruzzi G., Thalenfeld B.E., Tzagoloff A., Macino G.;
RT "Assembly of the mitochondrial membrane system. Structure and
RT nucleotide sequence of the gene coding for subunit 1 of yeast
RT cytochrome oxidase.";
RL J. Biol. Chem. 255:11927-11941(1980).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=91016898; PubMed=2216759;
RA Sarquell B., Hatait D., Delahodde A., Jacq C.;
RT "In vivo and in vitro analyses of an intron-encoded DNA endonuclease
RT from yeast mitochondria. Recognition site by site-directed
RT mutagenesis.";
RL Nucleic Acids Res. 18:5659-5665(1990).
CC -|- FUNCTION: ENDONUCLEASE INVOLVED IN INTRON HOMING. INTRODUCES A
CC SPECIFIC DOUBLE-STRAND BREAK AT THE JUNCTION OF THE TWO EXONS AA-
CC AS OF THE EXON1 GENE AND THUS MEDIATES THE INSERTION OF THE INTRON
CC INTO AN INTRONLESS STRAIN. RECOGNIZES AND CLEAVES THE SEQUENCE
CC 5'-TTTGTATCTTTGGTCACCTGAAGTATA-3'.
CC -|- SIMILARITY: TO OTHER INTRON NUCLEASES.


```
CC PROCESS OF PHOTOSYNTHESIS.
CC -!- COFACTOR: Binds 4 bacteriochlorophylls, 4 magnesium ions, 2
CC bacterioopheophytins, 2 ubiquinones, and 1 iron ion per trimer.
CC -!- SURUNIT: Heterotrimer composed of subunits L, M, and H.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M18412; AAA26176.1; -.
DR EMBL; X63378; CAA44976.1; -.
DR EMBL; AF195122; AAF24269.1; -.
DR PIR; A26538; A26538.
DR PIR; T50725; T50725.
DR PDB; 2RCR; 15-JUL-93.
DR PDB; 4RCR; 31-OCT-93.
DR PDB; 1PSS; 30-APR-94.
DR PDB; 1PST; 30-APR-94.
DR PDB; 1PCR; 07-FEB-95.
DR PDB; 1AIG; 22-OCT-97.
DR PDB; 1AIJ; 22-OCT-97.
DR PDB; 1YST; 27-FEB-95.
DR PDB; 1MPS; 14-OCT-98.
DR PDB; 1DS8; 08-MAR-00.
DR PDB; 1DV3; 08-MAR-00.
DR PDB; 1DV6; 08-MAR-00.
DR PDB; 1E14; 23-AUG-00.
DR PDB; 1B6D; 30-OCT-00.
DR PDB; 1F6N; 18-APR-01.
DR PDB; 1FNP; 18-APR-01.
DR PDB; 1FNQ; 18-APR-01.
DR PDB; 1JGW; 05-SEP-01.
DR PDB; 1JGX; 05-SEP-01.
DR PDB; 1JGY; 05-SEP-01.
DR PDB; 1JGZ; 05-SEP-01.
DR PDB; 1JHO; 05-SEP-01.
DR PDB; 1KBY; 13-NOV-02.
DR PDB; 1L9B; 12-JUN-02.
DR PDB; 1L9J; 12-JUN-02.
DR PDB; 1M3X; 28-AUG-02.
DR PDB; 1OOV; 13-DEC-99.
DR InterPro; IPR005652; PRCH.
DR Pfam; PF05239; PRCH; 1.
DR Pfam; PF03967; PRCH; 1.
DR TIGRFAMs; TIGR01150; pubA; 1.
DR Electron transport; Photosynthesis; Reaction center;
DR Bacteriochlorophyll; Transmembrane; 3D-structure.
FT DOMAIN 1 11 PERIPLASMIC.
FT TRANSMEM 12 31
FT DOMAIN 32 260 CYTOPLASMIC.
FT CONFLICT 38 38 E -> Q (IN REF. 3).
FT CONFLICT 81 81 E -> K (IN REF. 3).
FT CONFLICT 107 107 D -> N (IN REF. 3).
FT HELIX 12 34
FT TURN 35 39
FT STRAND 43 43
FT TURN 45 46
FT STRAND 49 49
FT STRAND 62 65
FT TURN 67 69
FT STRAND 72 75
FT STRAND 87 89
FT TURN 94 95
FT STRAND 98 100
FT HELIX 104 107
FT TURN 108 108
FT HELIX 110 112
FT STRAND 123 123
FT TURN 125 126

us-09-973-025-74.oli.rsp

FT STRAND 129 129
FT STRAND 131 133
FT TURN 134 136
FT TURN 138 139
FT STRAND 141 144
FT TURN 149 150
FT STRAND 152 154
FT TURN 156 157
FT STRAND 160 170
FT TURN 171 174
FT STRAND 175 182
FT TURN 184 185
FT STRAND 188 192
FT HELIX 193 195
FT STRAND 197 198
FT STRAND 203 204
FT HELIX 210 215
FT TURN 222 223
FT HELIX 227 243
FT TURN 244 244
FT HELIX 245 247
FT HELIX 251 254
SQ SEQUENCE 260 AA; 28035 MW; 7E0B913B63D4DEB0 CRC64;

Query Match 28.6%; Score 6; DB 1; Length 260;
Best Local Similarity 100.0%; Pred.No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 LAGLIY 18
Db 24 LAGLIY 29

RESULT 10
UPK1_BACHD STANDARD; PRT; 274 AA.
ID Q9KFL5;
AC Q9KFL5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative undecaprenol kinase 1 (EC 2.7.1.66) (Bacitracin resistance
DE protein).
GN UPK1 OR BACAL OR BH0464.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -!- FUNCTION: Probably phosphorylates undecaprenol to undecaprenyl
CC phosphate. Confers resistance to bacitracin (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + undecaprenol = ADP + undecaprenyl
CC phosphate.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- MISCELLANEOUS: Bacitracin is thought to be involved in inhibition
CC of peptidoglycan synthesis by sequestering undecaprenyl
CC diphosphate reducing the pool of lipid carrier available.
CC -!- SIMILARITY: BELONGS TO THE UPK FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

CC -----
DR EMBL; AP001508; BAB04183.1; --
DR PIR; H83707; H83707
DR HAMAP; MF_01006; ; 1
DR InterPro; IPR003824; BacA.
DR Pfam; PF02673; BacA; 1.
KW Transferrase; Kinase; Antibiotic resistance; Transmembrane;
KW Complete proteome.
FT TRANSMEM 7 29 POTENTIAL.
FT TRANSMEM 49 71 POTENTIAL.
FT TRANSMEM 92 112 POTENTIAL.
FT TRANSMEM 122 140 POTENTIAL.
FT TRANSMEM 193 215 POTENTIAL.
FT TRANSMEM 225 247 POTENTIAL.
FT TRANSMEM 254 273 POTENTIAL.
SQ SEQUENCE 274 AA; 30607 MW; 2AE1578AF80AF883 CRC64;

Query Match 28.6%; Score 6; DB 1; Length 274;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 WLGLI 17
Db 263 WLGLI 268

RESULT 11
YLMD_BACSU
ID YLMD_BACSU STANDARD; PRT; 278 AA.
AC O31726;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein ylmD.
GN YLMD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst P., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Gutseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinis S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivalta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RL subtilis."
RL Nature 390:249-256(1997).

```

```

CC -----
CC -!- SIMILARITY! BELONGS TO THE UPF0124 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z99112; CAB13411.1; --
CC PIR; D69876; D69876.
CC Subtilisin; BG13373; Ylmd.
DR InterPro; IPR003730; DUF152.
DR Pfam; PF02578; DUF152; 1.
DR TIGRPFAMs; TIGR00726; TIGR00726; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 278 AA; 30944 MW; 334595679F37CF04 CRC64;

Query Match 28.6%; Score 6; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 ESLNTG 10
Db 47 ESLNTG 52

RESULT 12
THER_BACTH
ID THER_BACTH STANDARD; PRT; 316 AA.
AC P00800;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Thermolysin (EC 3.4.24.27).
OS Bacillus thermoproteolyticus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1427;
RN [1]
RP SEQUENCE.
RA Titani K., Hermanson M.A., Ericsson L.H., Walsh K.A., Neurath H.;
RA "Amino-acid sequence of thermolysin."
RL Nature New Biol. 238:35-37(1972).
RN [2]
RP ACTIVE SITE.
RX MEDLINE=74052951; PubMed=4808703;
RA Burstein Y., Walsh K.A., Neurath H.;
RA "Evidence of an essential histidine residue in thermolysin."
RL Biochemistry 13:205-210(1974).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RX MEDLINE=83085812; PubMed=7175940;
RA Holmes M.A., Matthews B.W.;
RA "Structure of thermolysin refined at 1.6-A resolution."
RL J. Mol. Biol. 160:623-639(1982).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=75041142; PubMed=4214815;
RA Matthews B.W., Weaver L.H., Kester W.R.;
RA "The conformation of thermolysin."
RL J. Biol. Chem. 249:8030-8044(1974).
RN [5]
RP STRUCTURE BY NMR OF 255-316.
RX MEDLINE=95086079; PubMed=7993910;
RA Rico M., Jimenez M.A., Gonzalez C., de Filippis V., Fontana A.;
RA "NMR solution structure of the C-terminal fragment 255-316 of
RT thermolysin: a dimer formed by subunits having the native
RT structure."
RL Biochemistry 33:14834-14847(1994).
RN [6]
RP STRUCTURE BY NMR OF 205-316.
RX MEDLINE=97452621; PubMed=9305992;

```


RA Conejero-Lara F., Gonzalez C., Jimenez M.A., Padmanabhan S.,
RA Mateo P.L., Rico M.;
RT "NMR solution structure of the 205-316 C-terminal fragment of
RT thermolysin. An example of dimerization coupled to partial
RT unfolding";
RL Biochemistry 36:11975-11983(1997).
CC -!- FUNCTION: THERMOSTABLE EXTRACELLULAR ZINC METALLOPROTEASE.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Xaa-[Leu > Xaa]-Phe.
CC -!- COFACTOR: BINDS 1 ZINC ION AND FOUR CALCIUM IONS WHICH
CC CONTRIBUTE TO THE UNUSUAL THERMAL STABILITY OF THIS ENDOPEPTIDASE.
CC TWO ARE CLOSE TOGETHER AND PARTICIPATE IN A NETWORK OF SALT LINKS
CC ALSO INVOLVING ASP-138, GLU-177, LYS-182, ASP-185, GLU-190, AND
CC ASP-191. A THIRD INTERACTS WITH ASP-57 AND A FOURTH WITH ASP-200.
CC GLU AT POSITION 302 APPEARS TO FORM A SALT LINK WITH LYS-262.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M4.
DR PIR; I40579; HYBST
DR PDB; 4TLN; 31-JAN-84.
DR PDB; 5TLN; 31-JAN-84.
DR PDB; 7TLN; 31-JAN-84.
DR PDB; 8TLN; 30-APR-94.
DR PDB; 1TMN; 09-JAN-89.
DR PDB; 2TMN; 15-JAN-91.
DR PDB; 3TMN; 09-JAN-89.
DR PDB; 4TMN; 09-JAN-89.
DR PDB; 5TMN; 09-JAN-89.
DR PDB; 6TMN; 09-JAN-89.
DR PDB; 1THL; 31-JAN-94.
DR PDB; 1TLP; 09-JAN-89.
DR PDB; 1HYT; 31-JUL-94.
DR PDB; 1TRL; 07-FEB-95.
DR PDB; 1LNA; 08-MAY-95.
DR PDB; 1LNB; 08-MAY-95.
DR PDB; 1LNC; 08-MAY-95.
DR PDB; 1LND; 08-MAY-95.
DR PDB; 1LNE; 08-MAY-95.
DR PDB; 1LNF; 08-MAY-95.
DR PDB; 1FJ3; 18-APR-01.
DR PDB; 1FJO; 18-APR-01.
DR PDB; 1FJQ; 18-APR-01.
DR PDB; 1FJT; 18-APR-01.
DR PDB; 1FUU; 18-APR-01.
DR PDB; 1FJV; 18-APR-01.
DR PDB; 1FJW; 18-APR-01.
DR PDB; 1GXW; 05-DEC-02.
DR PDB; 1KE1; 16-NOV-02.
DR PDB; 1KJO; 05-DEC-02.
DR PDB; 1KJP; 05-DEC-02.
DR PDB; 1KKK; 10-DEC-02.
DR PDB; 1KL6; 11-DEC-02.
DR PDB; 1KR6; 09-JAN-03.
DR PDB; 1KRO; 10-JAN-03.
DR PDB; 1KS7; 11-JAN-03.
DR PDB; 1KTO; 17-JAN-03.
DR PDB; 1L3F; 03-JUL-02.
DR PDB; 1OSO; 25-MAR-03.
DR PDB; 1QF0; 29-DEC-99.
DR PDB; 1QF1; 29-DEC-99.
DR PDB; 1QF2; 20-DEC-00.
DR PDB; 1TLI; 18-FEB-00.
DR PDB; 2TLI; 13-MAR-00.
DR PDB; 3TLI; 13-MAR-00.
DR PDB; 4TLI; 13-MAR-00.
DR PDB; 5TLI; 13-MAR-00.
DR PDB; 6TLI; 13-MAR-00.
DR PDB; 7TLI; 13-MAR-00.
DR PDB; 8TLI; 13-MAR-00.
DR MEROPS; M04.001; -.
DR InterPro; IPR001570; Peptidase M4.
DR InterPro; IPR006025; Zn_Mtpeptidase.
DR Pfam; PF01447; Peptidase M4; 1.
DR Pfam; PF02868; Peptidase M4 C; 1.
DR PRINTS; PR00730; THERMOLYSIN.

DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Calcium; 3D-structure.
FT METAL 142 142 ZINC (CATALYTIC).
FT ACT_SITE 143 143
FT METAL 146 146 ZINC (CATALYTIC).
FT METAL 166 166 ZINC (CATALYTIC).
FT ACT_SITE 231 231 PROTON DONOR.
FT STRAND 4 11
FT TURN 13 14
FT STRAND 17 24
FT STRAND 28 29
FT STRAND 31 32
FT TURN 33 34
FT TURN 36 37
FT STRAND 39 43
FT TURN 45 46
FT STRAND 53 54
FT STRAND 56 57
FT STRAND 61 62
FT HELIX 65 67
FT HELIX 68 88
FT TURN 92 94
FT STRAND 100 104
FT STRAND 113 115
FT STRAND 120 122
FT STRAND 130 130
FT HELIX 133 135
FT HELIX 137 151
FT TURN 152 153
FT HELIX 159 180
FT TURN 181 181
FT STRAND 187 188
FT TURN 190 192
FT STRAND 193 193
FT TURN 195 196
FT TURN 198 199
FT STRAND 203 204
FT HELIX 208 211
FT TURN 212 212
FT HELIX 217 219
FT HELIX 225 229
FT TURN 230 233
FT HELIX 234 246
FT STRAND 248 250
FT TURN 251 252
FT STRAND 253 255
FT HELIX 260 273
FT TURN 274 274
FT TURN 277 278
FT HELIX 281 296
FT TURN 298 299
FT HELIX 301 312
FT TURN 313 314
SQ SEQUENCE 316 AA; 34333 MW; 004EDAF478744BB4 CRC64;
Query Match 28.6%; Score 6; DB 1; Length 316;
Best Local Similarity 100.0%; Pred.No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 14 AGLIYQ 19
DB 153 AGLIYQ 158
RESULT 13
POLG HCVH8
ID POLG HCVH8 STANDARD; PRT; 321 AA.
AC P27956;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Matrix protein (Envelope protein M);
DE Major envelope protein E; Nonstructural protein NS1] (Fragment).

```
OS Hepatitis C virus (isolate HCT18) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111110;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins.";
RL Virology 180:842-848(1991).
CC -!- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X53131; CAA37291.1; -.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV env.
CC InterPro; IPR002531; HCV NS1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV env; 1.
CC Pfam; PF01560; HCV NS1; 1.
CC ProDom; PD186062; HCV NS1; 1.
CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;
CC Transmembrane; Nonstructural protein.
CC NON TER 1
CC CHAIN <1 76 MATRIX PROTEIN (POTENTIAL).
CC CHAIN 77 267 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
CC CHAIN 268 >321 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
CC CARBOHYD 80 80 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 93 93 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 118 118 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 189 189 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 307 307 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 314 314 N-LINKED (GLCNAC. .) (POTENTIAL).
CC NON TER 321 321
CC SEQUENCE 321 AA; 34238 MW; 2F5DE79F7C7845C8 CRC64;
Query Match 28.6%; Score 6; DB 1; Length 321;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 SLNTGW 11
DB 316 SLNTGW 321
RESULT 14
POLG_HCVTH STANDARD; PRT; 321 AA.
AC P27557;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Matrix protein (Envelope protein M);
DE Major envelope protein B; Nonstructural protein NS1] (Fragment).
OS Hepatitis C virus (isolate TH) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111117;
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins.";
RL Virology 180:842-848(1991).
CC -!- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X53134; CAA37294.1; -.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV env.
CC InterPro; IPR002531; HCV NS1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV env; 1.
CC Pfam; PF01560; HCV NS1; 1.
CC ProDom; PD186062; HCV NS1; 1.
CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;
CC Transmembrane; Nonstructural protein.
CC NON TER 1
CC CHAIN <1 75 MATRIX PROTEIN (POTENTIAL).
CC CHAIN 76 267 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
CC CHAIN 268 >321 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
CC CARBOHYD 80 80 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 93 93 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 118 118 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 189 189 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 307 307 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 314 314 N-LINKED (GLCNAC. .) (POTENTIAL).
CC NON TER 321 321
CC SEQUENCE 321 AA; 34074 MW; B2EB83F521C3B520 CRC64;
Query Match 28.6%; Score 6; DB 1; Length 321;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 06; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 SLNTGW 11
DB 316 SLNTGW 321
RESULT 15
MTN3_NEILA STANDARD; PRT; 334 AA.
AC P24582;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Modification methylase NlaIII (EC 2.1.1.72) (Adenine-specific
DE methyltransferase NlaIII) (M.NlaIII).
GN NLAIII.
OS Neisseria lactamica.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=486;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 23970 / NRCC 2118;
```

RX MEDLINE=91117164; PubMed=2277628;
RA Labbe D., Hoeltke H.J., Lau P.C.K.;
RT "Cloning and characterization of two tandemly arranged DNA
RT methyltransferase genes of Neisseria lactamica: an adenine-specific
RT M.NlaIII and a cytosine-type methylase.";
RL Mol. Gen. Genet. 224:101-110(1990).
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC CATG, CAUSES SPECIFIC METHYLATION ON A-2 ON BOTH STRANDS AND
CC PROTECTS THE DNA FROM CLEAVAGE BY THE NlaIII ENDONUCLEASE.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-57 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54485; CAA38356.1; -.
DR PIR; S12036; XYNHAL.
DR REBASE; 3468; M.NlaIII.
DR InterPro; IPR002294; D12N6_mtfrase.
DR InterPro; IPR002052; N6_Mtase.
DR Pfam; PF02086; MethyltransfD12; 1.
DR PRINTS; PR00505; D12N6MTFRASE.
DR PROSITE; PS00092; N6_MTASE; 1.
KW Transferase; Methyltransferase; Restriction system.
SQ SEQUENCE 334 AA; 38382 MW; 8BCF1A1A35F02E89 CRC64;

Query Match 28.6%; Score 6; DB 1; Length 334;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 GLIYQH 20
Db 93 GLIYQH 98
|||||
|||||

Search completed: November 21, 2003, 20:59:53
Job time : 6.0925 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:45:57 ; Search time 25.62 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-74
Perfect score: 21
Sequence: 1 LNCNLSLNTGWLGLYQHK 21

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	52.4	96	12 Q91ZN5	Q91ZN5 hepatitis c
2	11	52.4	116	12 Q81443	Q81443 hepatitis c
3	11	52.4	130	12 Q91ZN6	Q91ZN6 hepatitis c
4	11	52.4	130	12 Q91ZP3	Q91ZP3 hepatitis c
5	11	52.4	130	12 Q91ZQ1	Q91ZQ1 hepatitis c
6	11	52.4	130	12 Q91ZM5	Q91ZM5 hepatitis c
7	11	52.4	130	12 Q91ZN1	Q91ZN1 hepatitis c
8	11	52.4	130	12 Q91ZM7	Q91ZM7 hepatitis c
9	11	52.4	130	12 Q91ZN2	Q91ZN2 hepatitis c
10	11	52.4	130	12 Q91ZP5	Q91ZP5 hepatitis c
11	11	52.4	130	12 Q91ZQ4	Q91ZQ4 hepatitis c
12	11	52.4	130	12 Q91ZN6	Q91ZN6 hepatitis c
13	11	52.4	130	12 Q91ZN4	Q91ZN4 hepatitis c
14	11	52.4	130	12 Q91ZN2	Q91ZN2 hepatitis c
15	11	52.4	130	12 Q91ZP0	Q91ZP0 hepatitis c
16	11	52.4	130	12 Q91ZQ5	Q91ZQ5 hepatitis c

17	11	52.4	130	12 Q91ZN7	Q91ZN7 hepatitis c
18	11	52.4	130	12 Q91ZQ2	Q91ZQ2 hepatitis c
19	11	52.4	130	12 Q91ZN3	Q91ZN3 hepatitis c
20	11	52.4	130	12 Q91ZQ6	Q91ZQ6 hepatitis c
21	11	52.4	130	12 Q91ZN3	Q91ZN3 hepatitis c
22	11	52.4	130	12 Q91ZM1	Q91ZM1 hepatitis c
23	11	52.4	130	12 Q91ZQ0	Q91ZQ0 hepatitis c
24	11	52.4	130	12 Q91ZQ3	Q91ZQ3 hepatitis c
25	11	52.4	130	12 Q91ZP8	Q91ZP8 hepatitis c
26	11	52.4	130	12 Q91ZP1	Q91ZP1 hepatitis c
27	11	52.4	130	12 Q91ZP7	Q91ZP7 hepatitis c
28	11	52.4	130	12 Q91ZP9	Q91ZP9 hepatitis c
29	11	52.4	130	12 Q91ZN8	Q91ZN8 hepatitis c
30	11	52.4	130	12 Q91ZP6	Q91ZP6 hepatitis c
31	11	52.4	130	12 Q91ZN0	Q91ZN0 hepatitis c
32	11	52.4	130	12 Q91ZM0	Q91ZM0 hepatitis c
33	11	52.4	130	12 Q91ZM4	Q91ZM4 hepatitis c
34	11	52.4	130	12 Q91ZP4	Q91ZP4 hepatitis c
35	11	52.4	130	12 Q91ZN9	Q91ZN9 hepatitis c
36	11	52.4	130	12 Q91ZM9	Q91ZM9 hepatitis c
37	11	52.4	137	12 Q9YK71	Q9YK71 hepatitis c
38	11	52.4	137	12 Q9YK70	Q9YK70 hepatitis c
39	11	52.4	139	12 Q9YK74	Q9YK74 hepatitis c
40	11	52.4	139	12 Q9YK76	Q9YK76 hepatitis c
41	11	52.4	139	12 Q91Z57	Q91Z57 hepatitis c
42	11	52.4	139	12 Q9YK77	Q9YK77 hepatitis c
43	11	52.4	139	12 Q9YJN1	Q9YJN1 hepatitis c
44	11	52.4	139	12 Q9YK75	Q9YK75 hepatitis c
45	11	52.4	140	12 Q92000	Q92000 hepatitis c

ALIGNMENTS

RESULT 1

Q91ZN5 PRELIMINARY; PRT; 96 AA.
AC Q91ZN5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_taxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X186A-b;
RX MEDLINE=20173694; PubMed=10708420;
RA Ray S.C., Mao O., Lanford R.E., Bassett S., Laeyendecker O.,
RA Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
RT replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL; AF230437; AAF68705.1; -;
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; I.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
SQ SEQUENCE 96 AA; 10158 MW; 72211BCD78DB768 CRC64;

Query Match 52.4%; Score 11; DB 12; Length 96;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNLSLNTGWL 11
| | | | | | | | | |
Db 55 LNCNLSLNTGWL 65

```

RESULT 2
Q81443          PRELIMINARY;          PRT;    116 AA.
AC Q81443;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Unknown protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CG2-5;
RA Hall W.W., Suzuki H.;
RT "Genetic heterogeneity of Hepatitis C viruses in a population of HIV
  infected intravenous drug abusers (IVDA)s."
RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL: L19172; AAA45593.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 116
SQ SEQUENCE 116 AA; 12761 MW; 57019C06E738F739 CRC64;

Query Match          52.4%; Score 11; DB 12; Length 116;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LNCNLSLNTGW 11
DB 52 LNCNLSLNTGW 62

RESULT 3
Q912M6          PRELIMINARY;          PRT;    130 AA.
AC Q912M6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X187A-e;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
  Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
  replication in chimpanzees."
RL J. Virol. 74:3058-3066(2000).
DR EMBL; AF230446; AAF68714.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 130
SQ SEQUENCE 130 AA; 13899 MW; 7A833BEE13183D24 CRC64;

Query Match          52.4%; Score 11; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LNCNLSLNTGW 11
DB 52 LNCNLSLNTGW 62

RESULT 4
Q912P3          PRELIMINARY;          PRT;    130 AA.
AC Q912P3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X268A-ff;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
  Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
  replication in chimpanzees."
RL J. Virol. 74:3058-3066(2000).
DR EMBL; AF230429; AAF68697.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 130
SQ SEQUENCE 130 AA; 13905 MW; 963BCFA4A304356E CRC64;

Query Match          52.4%; Score 11; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LNCNLSLNTGW 11
DB 55 LNCNLSLNTGW 65

RESULT 5
Q912Q1          PRELIMINARY;          PRT;    130 AA.
AC Q912Q1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X174A-o;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
  Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
  replication in chimpanzees."
RL J. Virol. 74:3058-3066(2000).
DR EMBL; AF230421; AAF68689.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 130
SQ SEQUENCE 130 AA; 13899 MW; 7A833BEE13183D24 CRC64;

Query Match          52.4%; Score 11; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LNCNLSLNTGW 11
DB 55 LNCNLSLNTGW 65

```

```

FT NON TER 130 130
SQ SEQUENCE 130 AA; 13934 MW; CAD37EF418D93692 CRC64;

Query Match 52.4%; Score 11; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
   |||||
Db 55 LNCNESLNTGW 65

RESULT 6
Q91ZM5
ID Q91ZM5 PRELIMINARY; PRT; 130 AA.
AC Q91ZM5;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=X361A-a;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RA Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
RT replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL: AF230447; AAF68715.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON TER 130 130
SQ SEQUENCE 130 AA; 13915 MW; 6AD37EEE13042124 CRC64;

Query Match 52.4%; Score 11; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
   |||||
Db 55 LNCNESLNTGW 65

RESULT 7
Q91ZN1
ID Q91ZN1 PRELIMINARY; PRT; 130 AA.
AC Q91ZN1;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=X187HC4.1;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RA Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
RT replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).

```

```

DR EMBL: AF230441; AAF68709.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON TER 130 130
SQ SEQUENCE 130 AA; 13875 MW; 663BDEA4A3043573 CRC64;

Query Match 52.4%; Score 11; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
   |||||
Db 55 LNCNESLNTGW 65

RESULT 8
Q91ZM7
ID Q91ZM7 PRELIMINARY; PRT; 130 AA.
AC Q91ZM7;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=X187A-ss;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RA Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
RT replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL: AF230445; AAF68713.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON TER 130 130
SQ SEQUENCE 130 AA; 13916 MW; 3A79744413094129 CRC64;

Query Match 52.4%; Score 11; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNCNESLNTGW 11
   |||||
Db 55 LNCNESLNTGW 65

RESULT 9
Q91ZN2
ID Q91ZN2 PRELIMINARY; PRT; 130 AA.
AC Q91ZN2;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=X187HC3 11; PubMed=10708420;
RX MEDLINE=20173694; Lanford R.E., Bassett S., Laeyendecker O.,
RA Ray S.C., Mao Q., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
RT replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL: AF230440; AAF68708.1; -.
DR InterPro: IPR002531; HCV NS1.
DR Pfam: PF01560; HCV NS1; 1.
DR ProDom: PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 130
SQ SEQUENCE 130 AA; 13915 MW; 6AD37EEE13042124 CRC64;

Query Match 52.4%; Score 11; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LNCNLSLNTGW 11
DB 55 LNCNLSLNTGW 65

RESULT 10
Q91ZP5 PRELIMINARY; PRT; 130 AA.
AC Q91ZP5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=X268A-a;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RT "Hypervariable region 1 sequence stability during hepatitis C virus
RT replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL: AF230427; AAF68695.1; -.
DR InterPro: IPR002531; HCV NS1.
DR Pfam: PF01560; HCV NS1; 1.
DR ProDom: PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 130
SQ SEQUENCE 130 AA; 13915 MW; 6AD37EEE13042124 CRC64;

Query Match 52.4%; Score 11; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LNCNLSLNTGW 11
DB 55 LNCNLSLNTGW 65

RESULT 11
Q91ZQ4 PRELIMINARY; PRT; 130 AA.
AC Q91ZQ4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

```

```

DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=X007A-f;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RT "Hypervariable region 1 sequence stability during hepatitis C virus
RT replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL: AF230418; AAF68686.1; -.
DR InterPro: IPR002531; HCV NS1.
DR Pfam: PF01560; HCV NS1; 1.
DR ProDom: PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 130
SQ SEQUENCE 130 AA; 13869 MW; 6FD73EF4D302E7B4 CRC64;

Query Match 52.4%; Score 11; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LNCNLSLNTGW 11
DB 55 LNCNLSLNTGW 65

RESULT 12
Q91ZN6 PRELIMINARY; PRT; 130 AA.
AC Q91ZN6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=X186A-11;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RT "Hypervariable region 1 sequence stability during hepatitis C virus
RT replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL: AF230436; AAF68704.1; -.
DR InterPro: IPR002531; HCV NS1.
DR Pfam: PF01560; HCV NS1; 1.
DR ProDom: PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 130
SQ SEQUENCE 130 AA; 13915 MW; 6AD37EEE13042124 CRC64;

Query Match 52.4%; Score 11; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LNCNLSLNTGW 11
DB 55 LNCNLSLNTGW 65

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:22 ; Search time 32.1 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-75
Perfect score: 20
Sequence: 1 AGLYQHKFNSSGCPERLAS 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

- 1: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2001.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2002.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	16 AAR84504	Hepatitis C virus
2	20	100.0	20	17 AAR90998	HCV E2 peptide E2
3	20	100.0	20	23 AAO18704	Hepatitis C virus
4	20	100.0	144	14 AAR33998	HC-J1 E2/NS1 prote
5	20	100.0	174	14 AAR34441	Sequence of glycop
6	20	100.0	319	15 AAR45330	Anti-HCV antibody
7	20	100.0	513	13 AAR24086	NANB hepatitis vir
8	20	100.0	527	13 AAR25136	HCV polypeptide 2.
9	20	100.0	733	14 AAR38278	NANB hepatitis vir

10	20	100.0	2894	13 AAR24440	Composite HCV HC-J
11	20	100.0	2894	16 AAR70230	Composite hepatitis
12	20	100.0	3011	15 AAR66995	Hepatitis C virus
13	16	80.0	20	24 ABP70587	Hepatitis C virus
14	16	80.0	30	23 AAU84627	HCV HepC segment
15	16	80.0	91	23 ABB77256	HCV bait polypepti
16	16	80.0	250	23 AAU79220	Hepatitis C virus
17	16	80.0	271	22 AAE00445	HCV E2-634 HVRI CO
18	16	80.0	271	22 AAE00446	HCV E2-634 HVRI-mu
19	16	80.0	350	23 AAE19891	Hepatitis C virus
20	16	80.0	363	24 AAE32880	Hepatitis C virus
21	16	80.0	363	24 ABP55567	Hepatitis C virus
22	16	80.0	497	22 AAE00443	HCV HVRI-7 constru
23	16	80.0	497	22 AAE00444	HCV HVRI-mut5 cons
24	16	80.0	1006	18 AAM12715	HCV genome type 1a
25	16	80.0	1997	23 AAU84802	HCV HepC cassette
26	16	80.0	2984	22 AAE00442	Hepatitis C virus
27	16	80.0	2984	22 AAE00447	Hepatitis C virus
28	16	80.0	2984	22 AAE00449	Hepatitis C virus
29	16	80.0	3011	13 AAR22154	NANBV Hutch c59 is
30	16	80.0	3011	19 AAU77397	Hepatitis C virus
31	16	80.0	3011	20 AAU98020	Infectious hepatitis
32	16	80.0	3011	22 AAB31169	Amino acid sequenc
33	16	80.0	3011	22 AAB59173	Protein encoded by
34	16	80.0	3011	23 AAU79221	Hepatitis C virus
35	16	80.0	3011	23 AAE19888	Hepatitis C virus
36	16	80.0	3011	23 AAU84597	HCV polypeptide 1a
37	16	80.0	3011	24 ABP71460	Amino acid sequenc
38	16	80.0	3012	23 AAU99289	Hepatitis C virus
39	16	80.0	3180	24 ABG73195	MKO-2 viral polype
40	16	80.0	3835	22 AAB31167	Amino acid sequenc
41	16	80.0	5985	23 AAU84799	HCV HepC1a scrambl
42	14	70.0	28	14 AAU34059	HCV E2 domain pept
43	14	70.0	38	20 AAV39502	HCV E2 protein res
44	14	70.0	38	20 AAY14177	HCV envelope regio
45	14	70.0	40	13 AAR20760	Peptide 10 based o

ALIGNMENTS

RESULT 1
AAR84504
ID AAR84504 standard; peptide; 20 AA.
AC AAR84504;
XX
DT 06-JAN-1997 (first entry)
XX
Hepatitis C virus peptide NS1-1* (residues 439-458).
XX
Hepatitis C virus; HCV; immunogen; non-structural region; NS1;
KW immunodominant; T cell epitope; vaccine.
XX
Hepatitis C virus.
OS
PN WO9513677-A2.
XX
PD 11-MAY-1995.
XX
PF 28-OCT-1994; 94WO-EP03555.
XX
PR 04-NOV-1993; 93EP-0402718.
XX
(INNO-) INNOGENETICS NV.
XX
Deleys R, Leroux-Roels G, Maertens G;
XX
WPI; 1995-193822/25.
XX
Hepatitis C Virus immunogenic polypeptide contg. a T-cell
PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in
PT production of vaccines, therapeutic agents, etc.

XX Example 4; Page 51; 105pp; English.

XX A series of overlapping peptides (including the present sequence) was

CC synthesised based on sequences in the core, E1 and E2/NS1 regions of

CC hepatitis C virus. The peptides were used as antigens in lympho-

CC proliferative assays to identify the main T-cell epitopes.

XX Sequence 20 AA;

SQ Query Match 100.0%; Score 20; DB 16; Length 20;

Best Local Similarity 100.0%; Pred. No. 9e-15;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCPERLAS 20

Db 1 AGLIYQHKFNSSGCPERLAS 20

RESULT 2

AAR90998

ID AAR90998 standard; peptide; 20 AA.

XX AC AAR90998;

XX DT 25-SEP-1996 (first entry)

XX DE HCV E2 peptide E2-\$1B for competition studies.

XX KW HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;

XX KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.

XX OS Synthetic.

XX PN WO9604385-A2.

XX PD 15-FEB-1996.

XX PF 31-JUL-1995; 95WO-EP03031.

XX PR 29-JUL-1994; 94EP-0870132.

XX PA (INNO-) INNOGENETICS NV.

XX PI Bosman F, Buyse M, De Martynoff G, Maertens G;

XX WPI; 1996-129401/13.

DR Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope

PT proteins - in presence of disulphide bond cleavage agent, to

PT produce proteins suitable for direct use in vaccines or diagnostic

PT assays of HCV

XX Example 7; Page 67; 146pp; English.

XX AAR90965-R90980 and AAR90995-R91015 represent synthetic hepatitis C

CC virus (HCV) E1 and E2 peptides used in competition studies. This

CC sequence represents a synthetic E2 peptide, and corresponds to residues

CC 439-458 of the E2 protein sequence. These sequences are useful for in

CC viro monitoring of HCV disease, or prognosis of the response to

CC interferon treatment of patients suffering from HCV infection. These

CC sequences compete with the proteins produced by AAT12704-T12709 and

CC AAT12961-T12974, which are included in vectors for the production of

CC recombinant E1, E2, and E1/E2. The recombinant proteins can then be

CC isolated and purified by carrying out a disulphide bond cleavage, or a

CC reduction step with a disulphide bond cleavage agent, after lysis of

CC recombinant host cells. The constructs containing the purified HCV

CC envelope proteins can be used for vaccinating humans against HCV, for in

CC viro detection of HCV antibodies in a sample, and in a serotyping assay

CC for detecting one or more serological types of HCV present in a

CC biological sample. The constructs can also be immobilised on a solid

CC substrate and incorporated into a reversed phase hybridisation assay for

CC determining the presence or the genotype of HCV. The new purification

CC method preserves the conformation of the recombinantly expressed E1, E2

CC and E1/E2, and eliminates contaminating proteins. Antigens isolated

CC using this method are more reactive with human sera than those isolated

CC by known techniques.

XX Sequence 20 AA;

SQ Query Match 100.0%; Score 20; DB 17; Length 20;

Best Local Similarity 100.0%; Pred. No. 9e-15;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCPERLAS 20

Db 1 AGLIYQHKFNSSGCPERLAS 20

RESULT 3

AAO18704

ID AAO18704 standard; Peptide; 20 AA.

XX AC AAO18704;

XX DT 24-OCT-2002 (first entry)

XX DB Hepatitis C virus E2 protein derived peptide E2-\$1B.

XX KW Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide;

XX KW immunostimulant; vaccine.

XX OS Hepatitis C virus.

XX PN WO200255548-A2.

XX PD 18-JUL-2002.

XX PF 11-JAN-2002; 2002WO-EP00219.

XX PR 11-JAN-2001; 2001US-260699P.

XX PR 30-AUG-2001; 2001US-315768P.

XX PA (INNO-) INNOGENETICS NV.

XX PI Maertens G, Bosman F, Buyse M;

XX WPI; 2002-599657/64.

XX New therapeutic vaccine compositions comprising at least one purified

PT recombinant hepatitis C virus (HCV) single or specific oligomeric

PT recombinant envelope protein E1 or E2, useful for immunizing humans

PT from HCV infection

XX Example 7; Page 226; 243pp; English.

XX The present invention relates to new therapeutic vaccine compositions for

CC inducing hepatitis C virus (HCV)-specific antibodies, comprising a

CC composition containing at least one purified recombinant HCV single or

CC specific oligomeric recombinant envelope proteins selected from an E1 and

CC an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are

CC useful for inducing HCV-specific antibodies or for immunising humans

CC against HCV. The recombinant HCV E1 and/or E2 proteins are useful as

CC vaccines or therapeutics, in HCV screening and confirmatory antibody

CC tests, for raising antibodies, in the preparation of medicament, and for

CC in vitro monitoring of HCV disease or prognosing the response to

CC treatment of patients suffering from HCV infection. The present sequence

CC is a peptide derived from the proteins of the invention.

XX Sequence 20 AA;

SQ Query Match 100.0%; Score 20; DB 23; Length 20;

Best Local Similarity 100.0%; Pred. No. 9e-15;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCPERLAS 20

Db 1 AGLIYQHKFNSSGCGPERLAS 20
|||||

RESULT 4

AAR33998
ID AAR33998 standard; Protein; 144 AA.

AC AAR33998;
DT 25-MAR-2003 (updated)
DT 26-JUL-1993 (first entry)
DE HC-J1 E2/NS1 protein.
KW Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
KW domain; immunological; cross-reactive; envelope protein; vaccine;
KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.
OS Synthetic.

PN 09306126-A1.
PD 01-APR-1993.
PF 11-SEP-1992; 92WO-US07683.
PR 13-SEP-1991; 91US-0759575.
PA (CHIR) CHIRON CORP.
PI Houghton M, Weiner AJ;
DR WPI; 1993-117468/14.
PT Immuno-reactive hepatitis C virus polypeptide compens. - contg.
PT at least 2 sequences from the first variable domain of distinct
PT HCV isolates
XX Disclosure; Fig 3; 106pp; English.

The sequences given in AAR33992-002 represent a portion of the E2/NS1 protein encoded by group I and group II HCV isolates, from amino acid 370-850. E2/NS1 are viral envelope proteins and are of immunogenic interest. E2/NS1 contains an N-terminal hypervariable domain of about 30 amino acids which shows large variation between nearly all isolates. This is an important immunoreactive domain. This putative envelope glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera virus) envelope polypeptide of the pestiviruses and the NS1 of the flaviviruses, both of which confer protective immunity in hosts vaccinated with these polypeptides. It has been discovered that a number of important HCV epitopes vary among viral isolates and that these epitopes can be mapped to specific domains. This meant that immunologically cross-reactive polypeptides which focus on variable rather than constant domains can be produced. See also AAQ39134-48 and AAR33982-91.
CC (Updated on 25-MAR-2003 to correct PN field.)

SQ Sequence 144 AA;

Query Match 100.0%; Score 20; DB 14; Length 144;
Best Local Similarity 100.0%; Pred. No. 5.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCGPERLAS 20
|||||
Db 70 AGLIYQHKFNSSGCGPERLAS 89

RESULT 5

AAR34441
ID AAR34441 standard; Protein; 174 AA.

XX AAR34441;
XX 25-MAR-2003 (updated)
DT 09-AUG-1993 (first entry)
XX Sequence of glycoprotein E2/NS1 in clone J1(JM).
DE Hepatitis C virus; envelope protein; glycoprotein; E2/NS1;
KW diagnostic reagent.
KW Hepatitis C virus.
OS Hepatitis C virus.
PN EP537626-A1.
PD 21-APR-1993.
PF 08-OCT-1992; 92EP-0117191.
PR 08-OCT-1991; 91JP-0260824.
PA (NAHE-) NAT INST OF HEALTH.

PI Harada S, Honda Y, Miyamura T, Saito I;
XX WPI; 1993-127516/16.
DR N-PSDB; AAQ40333.
XX Diagnostic reagent for hepatitis C virus - comprises second
PT envelope protein or first non-structural protein encoded by HCV
PT gene and has sugar chain
XX Claim 2; Pages 40-41; 58pp; English.

Glycoprotein E2/NS1 is derived from the second envelope protein or first non-structural protein encoded by the genome of HCV. The nucleic acid is extracted from the serum of the patient of hepatitis C. The serum is pref. mixed with transfer RNA (tRNA) as a carrier of virus RNA. As a technique of cloning cDNA from the nucleic acid, it is preferred to use polymerase chain reaction method. In the reaction, any commercially available random primers or synthesized DNA having a base sequence similar to that of primer AS1 may be used as a primer. Representative examples of sense primers include S1.
CC (Updated on 25-MAR-2003 to correct PN field.)

SQ Sequence 174 AA;

Query Match 100.0%; Score 20; DB 14; Length 174;
Best Local Similarity 100.0%; Pred. No. 6.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCGPERLAS 20
|||||
Db 100 AGLIYQHKFNSSGCGPERLAS 119

RESULT 6

AAR45330
ID AAR45330 standard; protein; 319 AA.

AC AAR45330;
XX 28-JUN-1994 (first entry)

DT Anti-HCV antibody reactive protein #1.
DE Hepatitis C virus; HCV; envelope; region; anti-HCV; antibody; vaccine.
KW Synthetic.
OS Key
FH Location/Qualifiers
FT Peptide 39..69
FT /label= EPI

/note= "Claim 1"

FT JP05310786-A.
 PN 22-NOV-1993.
 PD 30-APR-1992; 92JP-0111846.
 XX 30-APR-1992; 92JP-0111846.
 XX (OLYU) OLYMPUS OPTICAL CO LTD.
 XX WPI; 1994-002315/01.
 XX Peptide of HCV envelope region - having high reactivity against
 PT anti-hepatitis C virus antibody and having at least 6 amino acids
 XX Disclosure; Page 6-7; 11pp; Japanese.
 XX The sequences given in AAR45330-31 represent fragments derived from
 CC the hepatitis C virus (HCV) envelope region that are highly reactive
 CC against anti-HCV antibodies. These protein fragments may be used in
 CC vaccines against HCV. They have a high reactivity against anti-HCV
 CC antibodies regardless of the original HCV. They can be used to
 CC detect HCV infection but not the onset of infection.
 XX SQ Sequence 319 AA;
 Query Match 100.0%; Score 20; DB 15; Length 319;
 Best Local Similarity 100.0%; Pred. No. 1.1e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 AGLIYQHKFNSSGCPERLAS 20
 Db 248 AGLIYQHKFNSSGCPERLAS 267
 RESULT 7
 AAR24086
 ID AAR24086 standard; Protein; 513 AA.
 AC AAR24086;
 XX 09-NOV-1992 (first entry)
 DT NABH hepatitis virus strain HC-J1 gene product.
 DE non-A, non-B hepatitis virus; NANBH; HC-J5; PCR; amplification
 XX polymerase chain reaction; vaccine; antibody.
 XX Non-A, non-B hepatitis virus.
 XX EP485209-A.
 PN 13-MAY-1992.
 PD 07-NOV-1991; 91EP-0310297.
 XX 08-NOV-1990; 90JP-0304405.
 XX (IMMO) IMMUNO JAPAN INC.
 XX Nakamura T, Okamoto H;
 XX WPI; 1992-160959/20.
 DR N-PSDB; AAQ24466.
 XX Recombinant cDNA of NANBH virus strain HC-J5 and corresp.
 PT peptides - useful for diagnosis and in vaccines and immunological
 PT pharmaceuticals
 XX Disclosure; Page 19; 42pp; English.

CC This sequence is the gene product from the genome of the non-A, non-B
 CC hepatitis virus (NANBH) strain HC-J1. This sequence was derived
 CC from the DNA sequence which was amplified by polymerase chain reaction.
 CC The nucleotide sequences derived from this amplification can be used to
 CC detect NANBH infection which could not be detected by conventional
 CC methods. The detection kits allow highly specific and sensitive
 CC detection at an early phase of infection. This polypeptide sequence
 CC can be used for the manufacture of vaccines and immunological
 CC pharmaceuticals and also to produce antibodies specific to NANBH.
 XX SQ Sequence 513 AA;
 Query Match 100.0%; Score 20; DB 13; Length 513;
 Best Local Similarity 100.0%; Pred. No. 1.7e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCPERLAS 20
 Db 435 AGLIYQHKFNSSGCPERLAS 458

RESULT 8
 AAR25136
 ID AAR25136 standard; Protein; 527 AA.

AC AAR25136;
 XX 23-DEC-1992 (first entry)
 DT HCV polypeptide 2.
 DE Hepatitis C virus; blood transfusion.
 XX Synthetic.
 OS JP04159298-A.
 PN 02-JUN-1992.
 PD 19-OCT-1990; 90JP-0282431.
 PF 19-OCT-1990; 90JP-0282431.
 XX (OLYU) OLYMPUS OPTICAL CO LTD.
 XX WPI; 1992-231947/28.
 XX New peptides acting as antigenic analogues of human hepatitis C
 PT virus - useful for detecting HCV antibody positive patients
 XX Disclosure; Fig 2; 14pp; Japanese.

XX The sequences given in AAR25135-36 are peptides from the hepatitis C
 CC virus (HCV) which are recognised by the peptide sequences given in
 CC AAR25130-24. These antigenic peptides can be used on their own or as
 CC a mixture two different peptides. Using these peptides, HCV antibody
 CC positive patients can be detected and hepatitis caused by blood
 CC transfusion can be prevented.

XX SQ Sequence 527 AA;
 Query Match 100.0%; Score 20; DB 13; Length 527;
 Best Local Similarity 100.0%; Pred. No. 1.7e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 AGLIYQHKFNSSGCPERLAS 20
 Db 453 AGLIYQHKFNSSGCPERLAS 472

RESULT 9
 AAR38278
 ID AAR38278 standard; Protein; 733 AA.

```

XX AAR38278;
XX AC
XX DT 21-OCT-1993 (first entry)
XX DE NAMB hepatitis virus HC-OM gene polypeptide P-733-1.
XX XX
XX KW Non-A, non-B; virus; polymerase chain reaction; detection;
XX KW sensitive; specific; HCV; NANBH.
XX XX
XX OS Non-A, non-B hepatitis virus.
XX XX
XX FN JP05091884-A.
XX XX
XX PD 16-APR-1993.
XX XX
XX PP 10-APR-1991; 91JP-0196175.
XX XX
XX PR 12-JUN-1990; 90JP-0153401.
XX PR 08-NOV-1990; 90JP-0304405.
XX XX
XX PA (NAKA/) NAKAMURA T.
XX XX
XX WP 1993-199637/25.
XX DR N-PSDB; AAQ43888.
XX XX
XX PT Antigen related to non-A and non-B hepatitis virus - comprises
XX PT non-translation region comprising 340 - 341 mols. of nucleotides,
XX PT non-translation region comprising 1885 - 2551 mols. of
XX PT nucleotides including region 1,149 and, etc.
XX XX
XX PS Claim 14; Page 32-35; 73pp; Japanese.
XX XX
XX CC The sequence is that of NAMB hepatitis virus HC-OM gene polypeptide
XX CC P-733-1. It may be used in a system for detecting NAMB hepatitis.
XX CC This method is highly specific and sensitive, and can detect NAMB
XX CC hepatitis virus which could not be detected by conventional methods.
XX XX
XX SQ Sequence 733 AA;

Query Match 100.0%; Score 20; DB 14; Length 733;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGLIYQHKFNSSGCPERLAS 20
Db 439 AGLIYQHKFNSSGCPERLAS 458
|||||
|||||

RESULT 10
AAR24440
ID AAR24440 standard; protein; 2894 AA.
XX AC
XX AC AAR24440;
XX DT
XX DT 25-MAR-2003 (updated)
XX DT 02-DEC-1992 (first entry)
XX DE Composite HCV HC-J1/CDC/CHI protein.
XX KW Hepatitis C virus; peptides; antibodies; ELISA.
XX OS Synthetic.
XX XX
XX FH Key Location/Qualifiers
XX FT Peptide 1..20
XX FT Peptide /label= 1
XX FT Peptide 7..26
XX FT Peptide /label= 2
XX FT Peptide 13..32
XX FT Peptide /label= 3
XX FT Peptide 37..56
XX FT Peptide /label= 4

```

```

FT Peptide 49..68
FT Peptide /label= 5
FT Peptide 61..80
FT Peptide /label= 6
FT Peptide 73..92
FT Peptide /label= 7
FT Peptide 1688..1707
FT Peptide /label= 8
FT Peptide 1694..1713
FT Peptide /label= 9
FT Peptide 1706..1725
FT Peptide /label= 10
FT Peptide 1712..1731
FT Peptide /label= 11
FT Peptide 1718..1737
FT Peptide /label= 12
FT Peptide 1724..1743
FT Peptide /label= 13
FT Peptide 1730..1749
FT Peptide /label= 14
FT Peptide 2263..2282
FT Peptide /label= 15
FT Peptide 2275..2294
FT Peptide /label= 16
FT Peptide 2287..2306
FT Peptide /label= 17
FT Peptide 2299..2318
FT Peptide /label= 18
FT Peptide 2311..2330
FT Peptide /label= 19
XX XX
XX EP489968-A1.
XX PN
XX PD 17-JUN-1992.
XX XX
XX PF 14-DEC-1990; 90EP-0124241.
XX PR 14-DEC-1990; 90EP-0124241.
XX XX
XX PA (INNO-) INNOGENETICS NV.
XX XX
XX PI Deleys RJ, Maertens G, Pollet D, Van Heuverswyn H;
XX XX WPI; 1992-201383/25.
XX DR
XX XX New synthetic peptides for detecting antibodies to hepatitis C
XX FT virus - useful in e.g. ELISA assays, and for detection of HCV
XX FT antigens or as immunogens
XX XX
XX PS Disclosure; Fig 1; 32pp; English.
XX XX
XX CC RNA viruses frequently exhibit a high rate of spontaneous mutation,
XX CC thus a virus is considered to be the same of equiv. to HCV if it
XX CC exhibits a global homology of more than 70 percent with the HCV HC-
XX CC J1/CDC/CHI composite sequence. The peptide fragments of this DNA
XX CC sequence indicated in the features table can immunologically mimic
XX CC proteins encoded by HCV. Additional amino acids or chemical gps.
XX CC may be added to either end of the peptides for the purpose of
XX CC creating a linker arm for attachment to a carrier. The peptides can
XX CC be used for the detection of antibodies specific for HCV. They may
XX CC be used in the form of kits, opt. with reagents such as
XX CC staphylococcal protein A, streptococcal protein G, avidin or
XX CC streptavidin. The peptides may also be used as immunogens for
XX CC raising antibodies.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX XX
XX SQ Sequence 2894 AA;

```

```

Query Match 100.0%; Score 20; DB 13; Length 2894;
Best Local Similarity 100.0%; Pred. No. 7.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 AGLIYQHKFNSSGCPERLAS 20

```

```

Db      439 AGLIYQHKFNSSGCPERLAS 458
|||||
RESULT 11
AAR70230
ID AAR70230 standard; protein; 2894 AA.
XX
AC AAR70230;
XX
DT 25-MAR-2003 (updated)
DT 07-NOV-1995 (first entry)
XX
DE Composite hepatitis C virus (HC-J1/CDC/CHI).
XX
KW Composite hepatitis C virus; HC-J1/CDC/CHI; HCV; non-A non-B;
KW synthetic antigens; blood screening.
XX
OS Hepatitis C virus.
XX
PN EP644202-A1.
XX
PD 22-MAR-1995.
XX
PF 14-DEC-1990; 94EP-0108611.
XX
PR 14-DEC-1990; 90EP-0124241.
PR 14-DEC-1990; 90EP-0108611.
XX
PA (INNO-) INNOGENETICS NV.
XX
PI Deleys RJ, Maertens G, Pollet D, Van Heuverswyn H;
XX WPI; 1995-116946/16.
XX
DR Synthetic antigens for the detection of hepatitis C virus
PT antibodies - comprise portions of the HCV peptide sequence, for
PT use in screening blood and blood products
XX
PS Disclosure; Fig 1; 51pp; English.
XX
CC AAR70230 is the composite hepatitis C virus (HC-J1/CDC/CHI) protein
CC from which the synthetic HCV antigens described in AAR70210-R70229
CC were derived. These synthetic antigens can be used to screen blood,
CC or blood products for the presence HCV; they can also be used in
CC various specific assays for the detection of HCV antibodies, and
CC antigens, or as immunogens.
CC (Updated on 25-MAR-2003 to correct PN field.)
CC (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 2894 AA;
Query Match 100.0%; Score 20; DB 16; Length 2894;
Best Local Similarity 100.0%; Pred. No. 7.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AGLIYQHKFNSSGCPERLAS 20
|||||
Db      439 AGLIYQHKFNSSGCPERLAS 458
|||||
RESULT 12
AAR66995
ID AAR66995 standard; Protein; 3011 AA.
XX
AC AAR66995;
XX
DT 01-AUG-1995 (first entry)
XX
DE Hepatitis C virus gene HC-J1/protein.
XX
KW Hepatitis C virus; HCV gene HC-J1/protein; specific antibodies.
XX

OS Hepatitis C virus.
XX
PN JP06284887-A.
XX
PD 11-OCT-1994.
XX
PF 10-DEC-1993; 93JP-0345753.
XX
PR 10-DEC-1992; 92JP-0360705.
XX
PA (IMMO ) IMMUNO JAPAN KK.
XX
DR WPI; 1994-362594/45.
DR N-PSDB; AAQ74770.
XX
HCV genes and the corresponding proteins - used in the production
PT of anti-HCV antibodies and the detection of HCV infection
XX
Claim 11; Pages 18-32; 35pp; Japanese.
XX
PS AAQ74770 encodes AAR66995 the HC-J1/protein, the cDNA can be used in
CC the construction of an expression vector for the transformation
CC of a host cell. The host cell can then be used in the production
CC of proteins and peptides, useful in the preparation of monoclonal
CC and polyclonal HCV-specific antibodies.
XX
SQ Sequence 3011 AA;
Query Match 100.0%; Score 20; DB 15; Length 3011;
Best Local Similarity 100.0%; Pred. No. 8.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AGLIYQHKFNSSGCPERLAS 20
|||||
Db      439 AGLIYQHKFNSSGCPERLAS 458
|||||
RESULT 13
ABP70587
ID ABP70587 standard; peptide; 20 AA.
XX
AC ABP70587;
XX
DT 22-APR-2003 (first entry)
XX
DE Hepatitis C virus E2 glycoprotein derived antigenic peptide.
XX
KW Flavivirus; vector; envelope protein; peptide delivery; immune response;
KW tumour-associated antigen; immunotherapy; cancer; HCV; E2 glycoprotein.
XX
OS Hepatitis C virus.
XX
PN WO2002102828-A2.
XX
PD 27-DEC-2002.
XX
PF 31-MAY-2002; 2002WO-US17374.
XX
PR 01-JUN-2001; 2001US-295265P.
XX
PA (ACAM-) ACAMBIS INC.
XX
PI Kleanthous H, Oros L, Miller C;
XX
DR WPI; 2003-167480/16.
XX
PT Novel flavivirus vector useful for delivering foreign peptides,
PT comprises an envelope protein that comprises a foreign peptide -
XX
PS Example; Page 18; 34pp; English.
XX
CC The specification describes a flavivirus vector, comprising an envelope
CC protein that comprises a foreign peptide. The vector is useful for

```


CC delivery of a peptide, e.g. an antigen to a patient. It can be
 CC administered to induce an immune response to a pathogen or tumour from
 CC which the antigen is derived. It also useful for delivering
 CC tumour-associated antigens for use in immunotherapeutic methods against
 CC cancer. Asp70571-96 represent B and T cell epitopes from viruses.
 CC They may be expressed using the vectors of the invention.
 XX

SQ Sequence 20 AA;

Query Match 80.0%; Score 16; DB 24; Length 20;
 Best Local Similarity 100.0%; Pred. No. 1.9e-10;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 YQHKFNSSGCPERLAS 20
 |||||
 Db 3 YQHKFNSSGCPERLAS 18

RESULT 14
 AAU84627
 ID AAU84627 standard; Peptide; 30 AA.

XX AAU84627;

DT 08-MAY-2002 (first entry)

DE HCV HepC1a segment 30.

KW Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
 KW viral infection; human immunodeficiency virus; melanoma;
 KW bacterial infection; Salmonella; Legionella; parasitic infection;
 KW Trypanosoma; Toxoplasma; Giardia.

OS Hepatitis C virus.

XX WO200190197-A1.

XX 29-NOV-2001.

PF 25-MAY-2001; 2001WO-AU00622.

PR 26-MAY-2000; 2000AU-0007761.

XX (AUSU) UNIV AUSTRALIAN NAT.

XX Thomson SA, Ramshaw IA;

XX WPI; 2002-147575/19.

DR N-PSDB; ABK36465.

XX New synthetic polypeptides having several different segments of at
 PT least one parent polypeptide linked together differently compared to
 PT the linkage in the parent polypeptide, for inducing immune response
 PT against a pathogen or cancer -

XX Example 2; Fig 26; 364pp; English.

XX The invention relates to a new synthetic polypeptide (I) comprising
 CC several different segments of at least one parent polypeptide linked
 CC together in a different relationship relative to their linkage in the
 CC parent polypeptide to impede, abrogate or otherwise alter at least one
 CC function associated with the parent polypeptide and for inducing an
 CC immune response against a pathogen or cancer. Also included are a
 CC synthetic polynucleotide encoding and a computer system for
 CC designing the synthetic polypeptides. The synthetic polypeptides and
 CC polynucleotides are referred to as a savine. The synthetic polypeptide is
 CC useful for modulating immune responses preferably directed against a
 CC pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix,
 CC colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone
 CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.
 CC Compositions comprising the polypeptide may be used in the treatment or
 CC prophylaxis against viral (such as infections caused by HIV (human
 CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis

CC virus, Epstein-Barr virus and respiratory syncytial virus); bacterial
 CC (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,
 CC Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic
 CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
 CC Trypanosoma, Toxoplasma and Giardia) infections. The present
 CC sequence is a peptide derived from a parent protein used to
 CC construct a savine of the invention.
 XX

SQ Sequence 30 AA;

Query Match 80.0%; Score 16; DB 23; Length 30;
 Best Local Similarity 100.0%; Pred. No. 2.7e-10;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 YQHKFNSSGCPERLAS 20
 |||||
 Db 10 YQHKFNSSGCPERLAS 25

RESULT 15
 ABB77256

ID ABB77256 standard; Protein; 91 AA.

XX ABB77256;

DT 28-JUN-2002 (first entry)

DE HCV bait polypeptide 4.

XX SID; selected interacting domain; HCV; hepatitis C virus;
 KW liver disease; liver cancer; virucide; hepatotropic; antiinflammatory;
 KW antibacterial.

OS Hepatitis C virus strain H77.

XX EP1178116-A1.

XX 06-FEB-2002.

PF 03-AUG-2000; 2000EP-0402225.

PR 03-AUG-2000; 2000EP-0402225.

XX (HYBR-) HYBRIGENICS SA.

XX Legrain P, Whiteside S, Wojcik J;

XX WPI; 2002-208115/27.

DR N-PSDB; ABL55588.

XX New selected interacting domain polypeptides and polynucleotides,
 PT useful for treating or preventing infections or pathologies caused by
 PT hepatitis C virus (HCV) or those linked to HCV infection -

XX Claim 26; SEQ ID 80; 61pp + sequence listing; English.

XX The invention relates to nucleic acids encoding polypeptides which are
 CC termed SID polypeptides (selected interacting domain). These polypeptides
 CC are the final products of a double selection method involving a first
 CC step of selection of Hepatitis C virus (HCV)-derived polynucleotides
 CC through a two-hybrid system, and a second selection step involving an
 CC alignment between the different polynucleotides selected at the first
 CC step. The activity of polypeptides of the invention may be described as,
 CC virucide, hepatotropic, antiinflammatory and antibacterial. The
 CC polypeptide, polynucleotide and compositions comprising them are useful
 CC for treating or preventing viral or a bacterial infection, specifically
 CC infections or pathologies caused by HCV, or those pathologies linked to
 CC HCV infection. These may include liver disease and liver cancer. The
 CC current sequence represents a HCV bait polypeptide.

XX Note: The sequence data for this patent is not represented in the
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.

SQ Sequence 91 AA;

Query Match 80.0%; Score 16; DB 23; Length 91;
Best Local Similarity 100.0%; Pred. No. 7.3e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 YQHKFNSSGCCPERLAS 20
|||||
Db 64 YQHKFNSSGCCPERLAS 79

Search completed: November 21, 2003, 20:58:03
Job time : 33.15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.6 Seconds
(without alignments)
79,832 Million cell updates/sec

Title: US-09-973-025-75

Perfect score: 20

Sequence: 1 AGLIYQHKFNSSGCPERLAS 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCUTS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-08-612-973-75
2	20	100.0	20	3	US-08-927-597-75
3	20	100.0	20	4	US-08-635-886C-31
4	20	100.0	25	1	US-08-262-037-91
5	20	100.0	31	1	US-08-262-037-92
6	20	100.0	38	1	US-08-262-037-93
7	20	100.0	52	1	US-08-262-037-94
8	20	100.0	144	1	US-08-440-103-20
9	20	100.0	144	1	US-08-440-542-20
10	20	100.0	144	1	US-08-231-368-20
11	20	100.0	144	1	US-08-440-210-20
12	20	100.0	144	4	US-09-046-604-20
13	20	100.0	174	1	US-08-460-806-19
14	20	100.0	174	1	US-08-325-630-19
15	20	100.0	180	2	US-08-483-695-40
16	20	100.0	180	2	US-07-965-285-40
17	20	100.0	180	2	US-08-487-211-40
18	20	100.0	180	3	US-09-201-912-40
19	20	100.0	2894	2	US-08-466-975A-23
20	20	100.0	2894	2	US-08-391-671A-23
21	20	100.0	2894	3	US-08-467-902A-23
22	20	100.0	2894	3	US-09-275-265-23
23	20	100.0	2894	4	US-09-941-611-23
24	16	80.0	3011	3	US-09-014-416-1
25	16	80.0	3012	3	US-08-811-566-2
26	16	80.0	3012	4	US-09-034-756-2
27	14	70.0	28	5	PCT-US92-07865-19

```

28 14 70.0 38 4 US-09-645-470-10 Sequence 10, Appl
29 14 70.0 40 1 US-08-262-037-10 Sequence 10, Appl
30 14 70.0 76 3 US-08-444-818-79 Sequence 79, Appl
31 14 70.0 130 3 US-08-444-818-81 Sequence 81, Appl
32 14 70.0 278 1 US-08-440-103-15 Sequence 15, Appl
33 14 70.0 278 1 US-08-440-542-15 Sequence 15, Appl
34 14 70.0 278 1 US-08-231-368-15 Sequence 15, Appl
35 14 70.0 278 1 US-08-440-210-15 Sequence 15, Appl
36 14 70.0 278 4 US-09-046-604-15 Sequence 15, Appl
37 14 70.0 305 3 US-08-478-073-2 Sequence 2, Appl
38 14 70.0 314 3 US-08-612-973-42 Sequence 42, Appl
39 14 70.0 314 3 US-08-927-597-42 Sequence 42, Appl
40 14 70.0 319 3 US-08-612-973-44 Sequence 44, Appl
41 14 70.0 319 3 US-08-927-597-44 Sequence 44, Appl
42 14 70.0 333 1 US-08-453-552-12 Sequence 12, Appl
43 14 70.0 333 2 US-08-710-637-12 Sequence 12, Appl
44 14 70.0 333 5 PCT-US93-00907-12 Sequence 12, Appl
45 14 70.0 337 1 US-08-188-281B-7 Sequence 7, Appl

```

ALIGNMENTS

```

RESULT 1 0
US-08-612-973-75
; Sequence 75, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-612-973-75

```

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 8e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGLIYQHKFNSSGCPERLAS 20
|||||

Db 1 AGLIYQHKFNSSGCPERLAS 20

RESULT 2

US-08-927-597-75
; Sequence 75, Application US/08927597
; Patent No. 6245503

; GENERAL INFORMATION:

; APPLICANT: MAERTENS, GEERT

; APPLICANT: BOSMAN, FONS

; APPLICANT: DE MARTYNOFF, GUY

; APPLICANT: BUYSE, MARIE-ANGE

; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

; NUMBER OF SEQUENCES: 111

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: NIXON & VANDERHYE P.C.

; STREET: 1100 NORTH GLEBE ROAD

; CITY: ARLINGTON

; STATE: VIRGINIA

; COUNTRY: U.S.A.

; ZIP: 22201-4714

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

; APPLICATION DATA:

; APPLICATION NUMBER: US/08/927.597

; FILING DATE:

; CLASSIFICATION:

; APPLICATION DATA:

; APPLICATION NUMBER: US 08/612.973

; FILING DATE: 11-MAR-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: BYRNE, THOMAS E.

; REGISTRATION NUMBER: 32.205

; REFERENCE/DOCKET NUMBER: 1487-10

; TELEPHONE: (703) 816-4000

; TELEFAX: (703) 816-4100

; INFORMATION FOR SEQ ID NO: 75:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 20 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-927-597-75

Query Match 100.0%; Score 20; DB 3; Length 20;

Best Local Similarity 100.0%; Pred. No. 8e-15;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCPERLAS 20

Db 1 AGLIYQHKFNSSGCPERLAS 20

RESULT 3

US-08-635-886C-31

; Sequence 31, Application US/08635886C

; Patent No. 6555114

; GENERAL INFORMATION:

; APPLICANT: LEROUX-ROELS, Geert

; APPLICANT: DELEYS, Robert

; APPLICANT: MAERTENS, Geert

; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C

; TITLE OF INVENTION: VIRUS

; FILE REFERENCE: 2752-18

; CURRENT APPLICATION NUMBER: US/08/635,886C

; CURRENT FILING DATE: 1996-04-25

; PRIOR APPLICATION NUMBER: PCT/EP94/03555

; PRIOR FILING DATE: 1994-10-28

; PRIOR APPLICATION NUMBER: EP 93402718.6

; PRIOR FILING DATE: 1993-11-04

; NUMBER OF SEQ ID NOS: 286

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 31

; LENGTH: 20

; TYPE: PRT

; ORGANISM: hepatitis C virus

US-08-635-886C-31

Query Match 100.0%; Score 20; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 8e-15;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCPERLAS 20

Db 1 AGLIYQHKFNSSGCPERLAS 20

RESULT 4

US-08-262-037-91

; Sequence 91, Application US/08262037

; Patent No. 5747239

; GENERAL INFORMATION:

; APPLICANT: Chang Yi Wang and Barbara Hosein

; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR

; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV

; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES

; NUMBER OF SEQUENCES: 136

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNEGAN

; STREET: 345 PARK AVE.

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10154

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: WORDPERFECT 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/262,037

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/719,819

; FILING DATE: 24-June-1991

; APPLICATION NUMBER: 07/667,275

; FILING DATE: 11-Mar-1991

; APPLICATION NUMBER: 07/651,735

; FILING DATE: 07-Feb-1991

; APPLICATION NUMBER: 07/558,799

; FILING DATE: 26-July-1990

; APPLICATION NUMBER: 07/510,153

; FILING DATE: 16-April-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Maria C. H. Lin

; REGISTRATION NUMBER: 29,323

; REFERENCE/DOCKET NUMBER: 1151-4043 US3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-758-4800

; TELEFAX: (212) 751-6849

; TELEX: 421792

; INFORMATION FOR SEQ ID NO: 91:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 25 amino acids

; TYPE: Amino acid

; STRANDEDNESS:

; TOPOLOGY: Unknown

US-08-262-037-91

```
Query Match      100.0%; Score 20; DB 1; Length 25;
Best Local Similarity 100.0%; Pred. No. 9.8e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGLIYQHKFNSSGCCPERLAS 20
       |||||
Db      5 AGLIYQHKFNSSGCCPERLAS 24
       |||||

RESULT 5
US-08-262-037-92
; Sequence 92, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-262-037-92

Query Match      100.0%; Score 20; DB 1; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGLIYQHKFNSSGCCPERLAS 20
       |||||
Db      11 AGLIYQHKFNSSGCCPERLAS 30
       |||||

RESULT 6
US-08-262-037-93
; Sequence 93, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-262-037-93

Query Match      100.0%; Score 20; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.4e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGLIYQHKFNSSGCCPERLAS 20
       |||||
Db      18 AGLIYQHKFNSSGCCPERLAS 37
       |||||

RESULT 7
US-08-262-037-94
; Sequence 94, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
```

; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262.037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-262-037-94

Query Match 100.0%; Score 20; DB 1; Length 52;
Best Local Similarity 100.0%; Pred. No. 1.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCGPERLAS 20
Db .32 AGLIYQHKFNSSGCGPERLAS 51

RESULT 8
US-08-440-103-20
; Sequence 20, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440.103
; FILING DATE: 12-MAY-1995

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-103-20
Query Match 100.0%; Score 20; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 4.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AGLIYQHKFNSSGCGPERLAS 20
Db 70 AGLIYQHKFNSSGCGPERLAS 89
RESULT 9
US-08-440-542-20
; Sequence 20, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,542
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-542-20

Query Match      100.0%; Score 20; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 4.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AGLIYQHKFNSSGCGPERLAS 20
Db      70 AGLIYQHKFNSSGCGPERLAS 89

RESULT 10
US-08-231-368-20
; Sequence 20, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440.210
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231.368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-210-20

Query Match      100.0%; Score 20; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 4.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AGLIYQHKFNSSGCGPERLAS 20
Db      70 AGLIYQHKFNSSGCGPERLAS 89

RESULT 12
US-09-046-604-20
; Sequence 20, Application US/09046604
; Patent No. 6303292
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/046.604
; FILING DATE:
; CLASSIFICATION:
```

```
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-046-604-20

Query Match 100.0%; Score 20; DB 4; Length 144;
Best Local Similarity 100.0%; Pred. No. 4.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCPERLAS 20
   |||||
Db 70 AGLIYQHKFNSSGCPERLAS 89

RESULT 13
US-08-460-806-19
; Sequence 19, Application US/08460806
; Patent No. 5747241
; GENERAL INFORMATION:
; APPLICANT: MIYAMURA, TATSUO
; APPLICANT: SAITO, IZUMU
; APPLICANT: HARADA, SHIZUKO
; APPLICANT: HONDA, YOSHIKAZU
; TITLE OF INVENTION: DIAGNOSTIC REAGENT FOR HEPATITIS C
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,630
; FILING DATE: 19-OCT-1994
; APPLICATION NUMBER: US 07/956,993
; FILING DATE: 06-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5747241man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 4667-001-0
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 174 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-325-630-19

Query Match 100.0%; Score 20; DB 1; Length 174;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCPERLAS 20
   |||||
Db 100 AGLIYQHKFNSSGCPERLAS 119

RESULT 15
US-08-483-695-40
```

```
;
; LENGTH: 174 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-460-806-19

Query Match 100.0%; Score 20; DB 1; Length 174;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCPERLAS 20
   |||||
Db 100 AGLIYQHKFNSSGCPERLAS 119

RESULT 14
US-08-325-630-19
; Sequence 19, Application US/08325630
; Patent No. 5750331
; GENERAL INFORMATION:
; APPLICANT: MIYAMURA, TATSUO
; APPLICANT: SAITO, IZUMU
; APPLICANT: HARADA, SHIZUKO
; APPLICANT: HONDA, YOSHIKAZU
; TITLE OF INVENTION: DIAGNOSTIC REAGENT FOR HEPATITIS C
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/325,630
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/956,993
; FILING DATE: 06-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5750331man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 4667-001-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 174 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-325-630-19

Query Match 100.0%; Score 20; DB 1; Length 174;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCPERLAS 20
   |||||
Db 100 AGLIYQHKFNSSGCPERLAS 119

RESULT 15
US-08-483-695-40
```


; Sequence 40, Application US/08483695
; Patent No. 5866139
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Krensdorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; TITLE OF INVENTION: Applications
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: 483,695
; APPLICATION NUMBER: US/08/483,695
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/965,285
; FILING DATE: 18-MAR-1993
; APPLICATION NUMBER: FR 91 06 882
; FILING DATE: 06-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 05286-0001-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 180 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-483-695-40

Query Match 100.0%; Score 20; DB 2; Length 180;
Best Local Similarity 100.0%; Pred. No. 5.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCPERLAS 20
Db 106 AGLIYQHKFNSSGCPERLAS 125

Search completed: November 21, 2003, 21:15:14
Job time : 10.6 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time 19.5 Seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-75

Perfect score: 20

Sequence: 1 AGLYOHKFNSSGCPERLAS 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	10	US-09-973-025-75
2	20	100.0	20	11	US-09-899-303-75
3	20	100.0	20	11	US-09-995-808-75
4	20	100.0	20	11	US-09-995-860-75
5	20	100.0	20	12	US-09-995-791-75
6	20	100.0	2894	10	US-09-941-611-23
7	20	100.0	2894	15	US-10-044-995-23
8	16	80.0	91	10	US-09-921-397-80
9	16	80.0	250	10	US-09-952-572-8
10	16	80.0	350	10	US-09-929-955-4
11	16	80.0	350	14	US-10-104-966-4
12	16	80.0	363	12	US-10-128-587A-97
13	16	80.0	363	15	US-10-128-590-97
14	16	80.0	3011	9	US-09-742-659-4
15	16	80.0	3011	10	US-09-952-572-9

16	16	80.0	3011	10	US-09-929-955-1	Sequence 1, Appli
17	16	80.0	3011	10	US-09-747-419-20	Sequence 20, Appli
18	16	80.0	3011	11	US-09-891-894-3	Sequence 3, Appli
19	16	80.0	3011	12	US-10-184-150-3	Sequence 3, Appli
20	16	80.0	3011	14	US-10-104-966-1	Sequence 1, Appli
21	16	80.0	3011	15	US-10-259-275-20	Sequence 20, Appli
22	16	80.0	3012	10	US-09-938-076-2	Sequence 2, Appli
23	16	80.0	3012	11	US-09-995-937-2	Sequence 2, Appli
24	16	80.0	3012	11	US-09-917-563-2	Sequence 10, Appli
25	14	70.0	38	16	US-10-318-200-10	Sequence 3, Appli
26	14	70.0	254	10	US-09-407-430-3	Sequence 3, Appli
27	14	70.0	290	12	US-10-128-587A-3	Sequence 3, Appli
28	14	70.0	290	15	US-10-128-590-3	Sequence 3, Appli
29	14	70.0	301	12	US-10-128-587A-5	Sequence 5, Appli
30	14	70.0	301	15	US-10-128-590-5	Sequence 5, Appli
31	14	70.0	314	10	US-09-973-025-42	Sequence 42, Appli
32	14	70.0	314	11	US-09-899-303-42	Sequence 42, Appli
33	14	70.0	314	11	US-09-995-808-42	Sequence 42, Appli
34	14	70.0	314	11	US-09-995-860-42	Sequence 42, Appli
35	14	70.0	314	12	US-09-995-791-42	Sequence 42, Appli
36	14	70.0	319	10	US-09-973-025-44	Sequence 44, Appli
37	14	70.0	319	11	US-09-899-303-44	Sequence 44, Appli
38	14	70.0	319	11	US-09-995-808-44	Sequence 44, Appli
39	14	70.0	319	11	US-09-995-860-44	Sequence 44, Appli
40	14	70.0	319	12	US-09-995-791-44	Sequence 44, Appli
41	14	70.0	338	10	US-09-973-025-38	Sequence 38, Appli
42	14	70.0	338	11	US-09-899-303-38	Sequence 38, Appli
43	14	70.0	338	11	US-09-995-808-38	Sequence 38, Appli
44	14	70.0	338	11	US-09-995-860-38	Sequence 38, Appli
45	14	70.0	338	12	US-09-995-791-38	Sequence 38, Appli

ALIGNMENTS

RESULT 1

US-09-973-025-75
; Sequence 75, Application US/09973025
; Publication No. US20020182706A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100

```

: INFORMATION FOR SEQ ID NO: 75:
:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 20 amino acids
:   TYPE: amino acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:
:   MOLECULE TYPE: peptide
:   SEQUENCE DESCRIPTION: SEQ ID NO: 75:
US-09-973-025-75

```

Query Match 100.0%; Score 20; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCPERLAS 20
|||
Db 1 AGLIYQHKFNSSGCPERLAS 20

RESULT 2
US-09-899-303-75
; Sequence 75, Application US/09899303
; Publication NO. US20030036110A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
;

ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON

STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.

REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000

```

;
;
; INFORMATION FOR SEQ ID NO: 75:
;     SEQUENCE CHARACTERISTICS:
;         LENGTH: 20 amino acids
;

```

```

;
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 75:
US-09-899-303-75

Query Watch          100.0%;   Score 20;   DB 11;   Length 20;
Best Local Similarity 100.0%;   Pred.No. 1.5e-14;
Matches 20;   Conservative 0;   Mismatches 0;   Indels 0;   Gap
QY      1  AGLIYOHKFNSSGCPERLAS 20
      |||||

```

;
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-75

Query Match 100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCGPERLAS 20
|||||
Db 1 AGLIYQHKFNSSGCGPERLAS 20

RESULT 6

US-09-941-611-23
; Sequence 23, Application US/09941611
; Patent No. US20020106640A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; POLLET, DIRK
; MAERTENS, GEERT
; VAN HEUVERSWUN, HUGO

TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
ANTIBODIES TO HEPATITIS C VIRUS

NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESSES:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/941,611
FILING DATE: 30-Aug-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: 1995-02-21
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990

ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELEPHONE: 7038164000
TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:

US-09-941-611-23

Query Match 100.0%; Score 20; DB 10; Length 2894;
Best Local Similarity 100.0%; Pred. No. 1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCGPERLAS 20

Db 439 AGLIYQHKFNSSGCGPERLAS 458
|||||

RESULT 7

US-10-044-995-23
; Sequence 23, Application US/10044995
; Publication No. US20030049685A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; POLLET, DIRK
; MAERTENS, GEERT
; VAN HEUVERSWUN, HUGO

TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
ANTIBODIES TO HEPATITIS C VIRUS

NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESSES:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/044,995
FILING DATE: 15-Jan-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: <Unknown>
APPLICATION NUMBER: US 07/920,286
FILING DATE: 14-OCT-1992
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990

ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELEPHONE: 7038164000
TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:

US-10-044-995-23

Query Match 100.0%; Score 20; DB 15; Length 2894;
Best Local Similarity 100.0%; Pred. No. 1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCGPERLAS 20
|||||
Db 439 AGLIYQHKFNSSGCGPERLAS 458

RESULT 8

US-09-921-397-80
; Sequence 80, Application US/09921397
; Patent No. US20020151484A1

```
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: Sd nucleic acids and polypeptides selected from a
;   pathogenic strain of the hepatitis C virus and
;   applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 80
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-80

Query Match      80.0%; Score 16; DB 10; Length 91;
Best Local Similarity 100.0%; Pred. No. 9.2e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 YQHKFNSSGCGPERLAS 20
Db      64 YQHKFNSSGCGPERLAS 79

RESULT 9
US-09-952-572-8
; Sequence 8, Application US/09952572
; Patent No. US20020119495A1
; GENERAL INFORMATION:
; APPLICANT: HAWAII BIOTECHNOLOGY GROUP, Inc.
; APPLICANT: NAKANO, Eileen
; APPLICANT: CLEMENTS, David
; APPLICANT: HUMPHREYS, Tom
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITION OF HEPATITIS C
; FILE REFERENCE: HAW101100
; CURRENT APPLICATION NUMBER: US/09/952,572
; CURRENT FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: US 60/230,927
; PRIOR FILING DATE: 2000-09-13
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-09-952-572-8

Query Match      80.0%; Score 16; DB 10; Length 250;
Best Local Similarity 100.0%; Pred. No. 2.2e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 YQHKFNSSGCGPERLAS 20
Db      32 YQHKFNSSGCGPERLAS 47

RESULT 10
US-09-929-955-4
; Sequence 4, Application US/09929955
; Patent No. US20020136740A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
;   METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.23AUS2
; CURRENT APPLICATION NUMBER: US/09/929,955
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
```

```
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/225,767
; PRIOR FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-09-929-955-4

Query Match      80.0%; Score 16; DB 10; Length 350;
Best Local Similarity 100.0%; Pred. No. 2.9e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 YQHKFNSSGCGPERLAS 20
Db      64 YQHKFNSSGCGPERLAS 79

RESULT 11
US-10-104-966-4
; Sequence 4, Application US/10104966
; Publication No. US20020155124A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
;   METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.23AUSC1
; CURRENT APPLICATION NUMBER: US/10/104,966
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-10-104-966-4

Query Match      80.0%; Score 16; DB 14; Length 350;
Best Local Similarity 100.0%; Pred. No. 2.9e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 YQHKFNSSGCGPERLAS 20
Db      64 YQHKFNSSGCGPERLAS 79

RESULT 12
US-10-128-587A-97
; Sequence 97, Application US/10128587A
; Publication No. US20030152940A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Constructs and methods for expression of recombinant HCV envelope
;   proteins
; FILE REFERENCE: 134 PCT
; CURRENT APPLICATION NUMBER: US/10/128,587A
; CURRENT FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
; LENGTH: 363
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: hepatitis C virus
US-10-128-587A-97

Query Match      80.0%; Score 16; DB 12; Length 363;
Best Local Similarity 100.0%; Pred. No. 3e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 YQHKFNSSGCGPERLAS 20
Db      60 YQHKFNSSGCGPERLAS 75

RESULT 13
US-10-128-590-97
; Sequence 97, Application US/10128590
; Publication No. US20030108561A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Core-glycosylated HCV envelope proteins
; FILE REFERENCE: 135 PCT
; CURRENT APPLICATION NUMBER: US/10/128,590
; PRIOR FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-128-590-97

Query Match      80.0%; Score 16; DB 15; Length 363;
Best Local Similarity 100.0%; Pred. No. 3e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 YQHKFNSSGCGPERLAS 20
Db      60 YQHKFNSSGCGPERLAS 75

RESULT 14
US-09-742-659-4
; Sequence 4, Application US/09742659
; Patent No. US20010034019A1
; GENERAL INFORMATION:
; APPLICANT: Hong, Zhi
; APPLICANT: Butkiewicz, Nancy J.
; APPLICANT: Zhong, Weidong
; APPLICANT: Ingravallo, Paul
; APPLICANT: Wright-Minogue, Jacquelyn
; APPLICANT: Lau, Johnson Y.
; APPLICANT: Lemon, Stanley M.
; TITLE OF INVENTION: Chimeric HCV/GBV-B viruses
; FILE REFERENCE: ID01116
; CURRENT APPLICATION NUMBER: US/09/742,659
; CURRENT FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: US 60/171,469
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-742-659-4

Query Match      80.0%; Score 16; DB 9; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 YQHKFNSSGCGPERLAS 20
```

```
Db      443 YQHKFNSSGCGPERLAS 458

RESULT 15
US-09-952-572-9
; Sequence 9, Application US/09952572
; Patent No. US20020119495A1
; GENERAL INFORMATION:
; APPLICANT: HAWAII BIOTECHNOLOGY GROUP, Inc.
; APPLICANT: NAKANO, Eileen
; APPLICANT: CLEMENTS, David
; APPLICANT: HUMPHREYS, Tom
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITION OF HEPATITIS C
; FILE REFERENCE: HAWBIO1100
; CURRENT APPLICATION NUMBER: US/09/952,572
; CURRENT FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: US 60/230,927
; PRIOR FILING DATE: 2000-09-13
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-09-952-572-9

Query Match      80.0%; Score 16; DB 10; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 YQHKFNSSGCGPERLAS 20
Db      443 YQHKFNSSGCGPERLAS 458

Search completed: November 21, 2003, 22:19:37
Job time : 19.5 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 163.1 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-75

Perfect score: 20

Sequence: 1 AGLYQHKFNSSGCPERLAS 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_AA_Main.*

- 1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/1/paa/US097A_COMB.pep.*
- 22: /cgn2_6/ptodata/1/paa/US097B_COMB.pep.*
- 23: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
- 24: /cgn2_6/ptodata/1/paa/US099A_COMB.pep.*
- 25: /cgn2_6/ptodata/1/paa/US099B_COMB.pep.*
- 26: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
- 27: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
- 28: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
- 29: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
- 30: /cgn2_6/ptodata/1/paa/US104_COMB.pep.*
- 31: /cgn2_6/ptodata/1/paa/US106_COMB.pep.*
- 32: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	ID	Description
1	20	100.0	20 13	US-08-974-685-31
				Sequence 31, Appl

2	20	100.0	20 13	US-08-974-690-31	Sequence 31, Appl
3	20	100.0	20 13	US-08-974-690A-31	Sequence 31, Appl
4	20	100.0	20 13	US-08-974-690B-31	Sequence 31, Appl
5	20	100.0	20 13	US-08-974-690C-31	Sequence 31, Appl
6	20	100.0	20 23	US-09-899-303-75	Sequence 75, Appl
7	20	100.0	20 23	US-09-899-303A-75	Sequence 75, Appl
8	20	100.0	20 25	US-09-973-025-75	Sequence 75, Appl
9	20	100.0	20 25	US-09-995-791-75	Sequence 75, Appl
10	20	100.0	20 25	US-09-995-808-75	Sequence 75, Appl
11	20	100.0	20 25	US-09-995-860-75	Sequence 75, Appl
12	20	100.0	20 26	US-10-020-510-75	Sequence 75, Appl
13	20	100.0	20 29	US-10-321-798-75	Sequence 75, Appl
14	20	100.0	25 8	US-08-475-482-91	Sequence 91, Appl
15	20	100.0	25 8	US-08-477-072-91	Sequence 91, Appl
16	20	100.0	25 8	US-08-477-582-91	Sequence 91, Appl
17	20	100.0	25 8	US-08-480-253-91	Sequence 91, Appl
18	20	100.0	31 8	US-08-475-482-92	Sequence 92, Appl
19	20	100.0	31 8	US-08-477-072-92	Sequence 92, Appl
20	20	100.0	31 8	US-08-477-582-92	Sequence 92, Appl
21	20	100.0	31 8	US-08-480-253-92	Sequence 92, Appl
22	20	100.0	38 8	US-08-475-482-93	Sequence 93, Appl
23	20	100.0	38 8	US-08-477-072-93	Sequence 93, Appl
24	20	100.0	38 8	US-08-477-582-93	Sequence 93, Appl
25	20	100.0	38 8	US-08-480-253-93	Sequence 93, Appl
26	20	100.0	40 3	US-07-667-275A-10	Sequence 10, Appl
27	20	100.0	52 8	US-08-475-482-94	Sequence 94, Appl
28	20	100.0	52 8	US-08-477-072-94	Sequence 94, Appl
29	20	100.0	52 8	US-08-477-582-94	Sequence 94, Appl
30	20	100.0	52 8	US-08-480-253-94	Sequence 94, Appl
31	20	100.0	144 8	US-08-471-498-20	Sequence 20, Appl
32	20	100.0	513 3	US-07-790-382A-30	Sequence 30, Appl
33	20	100.0	733 3	US-07-866-045-11	Sequence 11, Appl
34	20	100.0	2894 26	US-10-044-995-23	Sequence 23, Appl
35	16	80.0	20 1	PCT-US02-17374-25	Sequence 25, Appl
36	16	80.0	91 24	US-09-921-397-80	Sequence 80, Appl
37	16	80.0	250 1	PCT-US01-28767-8	Sequence 8, Appl
38	16	80.0	250 25	US-09-952-572-8	Sequence 4, Appl
39	16	80.0	350 21	US-09-705-547-4	Sequence 4, Appl
40	16	80.0	350 24	US-09-929-955-4	Sequence 4, Appl
41	16	80.0	350 27	US-10-104-966-4	Sequence 4, Appl
42	16	80.0	350 32	US-60-229-175-4	Sequence 4, Appl
43	16	80.0	363 27	US-10-128-587A-97	Sequence 97, Appl
44	16	80.0	363 27	US-10-128-590-97	Sequence 97, Appl
45	16	80.0	421 29	US-10-365-620-70	Sequence 70, Appl

ALIGNMENTS

RESULT 1

US-08-974-685-31
; Sequence 31, Application US/08974685
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; DELEYS, ROBERT
; MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; HEPATITIS C VIRUS STRUCTURAL PROTEINS. PROPHYLACTIC AND
; THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 181
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,685
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-08-974-685-31

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCPERLAS 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 AGLIYQHKFNSSGCPERLAS 20

RESULT 2

US-08-974-690-31
Sequence 31, Application US/08974690
GENERAL INFORMATION:

APPLICANT: LEROUIX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 166

CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P. C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/635,886
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

US-08-974-690-31

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCPERLAS 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 AGLIYQHKFNSSGCPERLAS 20

RESULT 3

US-08-974-690A-31
Sequence 31, Application US/08974690A
GENERAL INFORMATION:

APPLICANT: LEROUIX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
MAERTENS, GEERT

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 177

CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P. C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690A
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 31:

US-08-974-690A-31

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCPERLAS 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 AGLIYQHKFNSSGCPERLAS 20

RESULT 4

US-08-974-690B-31

Sequence 31, Application US/08974690B
GENERAL INFORMATION:

APPLICANT: LEROUIX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
MAERTENS, GEERT

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690B
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 31:

US-08-974-690B-31

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGLIYQHKFNSSGCPERLAS 20
Db 1 AGLIYQHKFNSSGCPERLAS 20

RESULT 5

US-08-974-690C-31
Sequence 31, Application US/08974690C
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELEYS, Robert
APPLICANT: MAERTENS, Geert
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
FILE REFERENCE: 2551-94
CURRENT APPLICATION NUMBER: US/08/974,690C
CURRENT FILING DATE: 1997-11-19
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28
PRIOR APPLICATION NUMBER: EP 93402718.6
PRIOR FILING DATE: 1993-11-04
NUMBER OF SEQ ID NOS: 286
SOFTWARE: Patent In version 3.1
SEQ ID NO 31
LENGTH: 20
TYPE: PRT
ORGANISM: hepatitis C virus
US-08-974-690C-31

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGLIYQHKFNSSGCPERLAS 20

Db 1 AGLIYQHKFNSSGCPERLAS 20

RESULT 6

US-09-899-303-75
Sequence 75, Application US/09899303
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 75:
US-09-899-303-75

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGLIYQHKFNSSGCPERLAS 20
Db 1 AGLIYQHKFNSSGCPERLAS 20

RESULT 7

US-09-899-303A-75
Sequence 75, Application US/09899303A
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.

STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/899,303A
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/612,973
FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 75:

SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 75:

US-09-899-303A-75

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCGPERLAS 20

Db 1 AGLIYQHKFNSSGCGPERLAS 20

RESULT 8

US-09-973-025-75

Sequence 75, Application US/09973025

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

BOSMAN, FONS

DE MARTYNOFF, GUY

BUYSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/973,025

FILING DATE: 10-Oct-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 75:
US-09-973-025-75

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCGPERLAS 20

Db 1 AGLIYQHKFNSSGCGPERLAS 20

RESULT 9

US-09-995-791-75

Sequence 75, Application US/09995791

GENERAL INFORMATION:

APPLICANT: Innogenetics N.V.

TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and

TITLE OF INVENTION: therapeutic use.

FILE REFERENCE: 2551-68

CURRENT APPLICATION NUMBER: US/09/995,791

CURRENT FILING DATE: 2001-11-29

NUMBER OF SEQ ID NOS: 122

SOFTWARE: PatentIn 3.1

SEQ ID NO 75

LENGTH: 20

TYPE: PRT

ORGANISM: Hepatitis C virus

US-09-995-791-75

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCGPERLAS 20

Db 1 AGLIYQHKFNSSGCGPERLAS 20

RESULT 10

US-09-995-808-75

Sequence 75, Application US/09995808

GENERAL INFORMATION:

APPLICANT: Innogenetics N.V.

TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and

TITLE OF INVENTION: therapeutic use.

FILE REFERENCE: 2551-70

CURRENT APPLICATION NUMBER: US/09/995,808

CURRENT FILING DATE: 2001-11-29

NUMBER OF SEQ ID NOS: 122

SOFTWARE: PatentIn 3.1

SEQ ID NO 75

LENGTH: 20

TYPE: PRT

ORGANISM: Hepatitis C virus

US-09-995-808-75

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCGPERLAS 20
|||||
Db 1 AGLIYQHKFNSSGCGPERLAS 20

RESULT 11
US-09-995-860-75
; Sequence 75, Application US/09995860
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 75
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-75

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCGPERLAS 20
|||||
Db 1 AGLIYQHKFNSSGCGPERLAS 20

RESULT 12
US-10-020-510-75
; Sequence 75, Application US/10020510
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-72
; CURRENT APPLICATION NUMBER: US/10/020,510
; CURRENT FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 75
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-020-510-75

Query Match 100.0%; Score 20; DB 26; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCGPERLAS 20
|||||
Db 1 AGLIYQHKFNSSGCGPERLAS 20

RESULT 13
US-10-321-798-75
; Sequence 75, Application US/10321798
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-93
; CURRENT APPLICATION NUMBER: US/10/321,798
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: 60/418,358
; PRIOR FILING DATE: 2002-10-16

; PRIOR APPLICATION NUMBER: 10/020,510
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 75
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-321-798-75

Query Match 100.0%; Score 20; DB 29; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCGPERLAS 20
|||||
Db 1 AGLIYQHKFNSSGCGPERLAS 20

RESULT 14
US-08-475-482-91
; Sequence 91, Application US/08475482
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosen
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,482
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 17-June-1994
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-475-482-91

Job time : 163.1 secs

Query Match 100.0%; Score 20; DB 8; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCPERLAS 20
| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5 AGLIYQHKFNSSGCPERLAS 24

RESULT 15
US-08-477-072-91
; Sequence 91, Application US/08477072
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,072
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 17-June-1994
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-477-072-91

Query Match 100.0%; Score 20; DB 8; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCPERLAS 20
| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5 AGLIYQHKFNSSGCPERLAS 24


```
; SEQ ID NO 466
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 30
US-10-296-734-466

Query Match      80.0%; Score 16; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 2e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 YQHKFNSSGCPERLAS 20
Db      10 YQHKFNSSGCPERLAS 25
|||||
|||||

RESULT 3
US-10-296-734-816
; Sequence 816, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 816
; LENGTH: 1997
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette C
US-10-296-734-816

Query Match      80.0%; Score 16; DB 6; Length 1997;
Best Local Similarity 100.0%; Pred. No. 3.3e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 YQHKFNSSGCPERLAS 20
Db      610 YQHKFNSSGCPERLAS 625
|||||
|||||

RESULT 4
US-10-296-734-406
; Sequence 406, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a consensus polyprotein
US-10-296-734-406

Query Match      80.0%; Score 16; DB 6; Length 3011;
Best Local Similarity 100.0%; Pred. No. 4.4e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 YQHKFNSSGCPERLAS 20
Db      443 YQHKFNSSGCPERLAS 458
|||||
|||||

RESULT 5
US-10-296-734-810
; Sequence 810, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 810
; LENGTH: 5985
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a savine
US-10-296-734-810

Query Match      80.0%; Score 16; DB 6; Length 5985;
Best Local Similarity 100.0%; Pred. No. 7e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 YQHKFNSSGCPERLAS 20
Db      4630 YQHKFNSSGCPERLAS 4645
|||||
|||||

RESULT 6
US-10-685-435-19
; Sequence 19, Application US/10685435
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Multi-mer peptides derived from Hepatitis C Virus
; TITLE OF INVENTION: envelope proteins for diagnostic use and vaccination
; TITLE OF INVENTION: purposes
; FILE REFERENCE: PCT98.75.HCV30
; CURRENT APPLICATION NUMBER: US/10/685,435
; CURRENT FILING DATE: 2003-10-16
; PRIOR APPLICATION NUMBER: US/09/566,266B
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 97870179.5
; PRIOR FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: patenting Ver. 2.1
; SEQ ID NO 19
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-685-435-19

Query Match      70.0%; Score 14; DB 6; Length 38;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      7 HKFNSSGCPERLAS 20
Db      16 HKFNSSGCPERLAS 29
|||||
|||||

RESULT 7
PCT-US03-19834-3
; Sequence 3, Application PC/TUS0319834
; GENERAL INFORMATION:
```



```
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; FILE REFERENCE: IOWA:045W0
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-3

Query Match 70.0%; Score 14; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 9e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 HKFNSSGCGPERLAS 20
Db 63 HKFNSSGCGPERLAS 76

RESULT 8
US-10-655-562-4
; Sequence 4, Application US/10655562
; GENERAL INFORMATION:
; APPLICANT: BISHOP-HURLEY, SHARON L.
; APPLICANT: SCHMIDT, FRANCIS J.
; APPLICANT: SMITH, ARNOLD L.
; TITLE OF INVENTION: PHAGE-DISPLAY PEPTIDES AS NOVEL ANTIMICROBIAL AGENTS
; FILE REFERENCE: UVM0:022US
; CURRENT APPLICATION NUMBER: US/10/655,562
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: 60,409,909
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-655-562-4

Query Match 70.0%; Score 14; DB 6; Length 333;
Best Local Similarity 100.0%; Pred. No. 9e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 HKFNSSGCGPERLAS 20
Db 63 HKFNSSGCGPERLAS 76

RESULT 9
PCT-US03-33610-4
; Sequence 4, Application PC/TUS0333610
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: SELBY, Mark
; APPLICANT: PALIARD, Xavier
; TITLE OF INVENTION: ACTIVATION OF HCV-SPECIFIC T-CELLS
; FILE REFERENCE: 2300-1612.60
; CURRENT APPLICATION NUMBER: PCT/US03/33610

; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; FILE REFERENCE: IOWA:045W0
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
PCT-US03-33610-4

Query Match 70.0%; Score 14; DB 1; Length 637;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 HKFNSSGCGPERLAS 20
Db 273 HKFNSSGCGPERLAS 286

RESULT 10
PCT-US03-19834-2
; Sequence 2, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; FILE REFERENCE: IOWA:045W0
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; OTHER INFORMATION: Description of Artificial Sequence: Peptide
PCT-US03-19834-2

Query Match 70.0%; Score 14; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.9e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 HKFNSSGCGPERLAS 20
Db 445 HKFNSSGCGPERLAS 458

RESULT 11
US-10-651-165-181
; Sequence 181, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 181
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-181

Query Match          60.0%; Score 12; DB 6; Length 450;
Best Local Similarity 100.0%; Pred. No. 9.8e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AGLIYQHKNSS 12
Db      439 AGLIYQHKNSS 450
|||||

RESULT 12
PCT-US03-20322-211
; Sequence 211, Application PC/TUS0320322
; GENERAL INFORMATION:
; APPLICANT: Dana-Farber Cancer Institute, Inc., et al.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; FILE REFERENCE: DFN-043PC
; CURRENT APPLICATION NUMBER: PCT/US03/20322
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 211
; LENGTH: 2280
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
PCT-US03-20322-211

Query Match          55.0%; Score 11; DB 1; Length 2280;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      7 HKFNSSGCP 17
Db      445 HKFNSSGCP 455
|||||

RESULT 13
US-10-296-734-468
; Sequence 468, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 468
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 31
US-10-296-734-468

Query Match          50.0%; Score 10; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      11 SSGCERLAS 20
Db      118 SSGCERLAS 20
|||||
```

```
Db      1 SSGCERLAS 10

RESULT 14
US-09-638-693-158
; Sequence 158, Application US/09638693
; GENERAL INFORMATION:
; APPLICANT: Innogenetics sa.
; TITLE OF INVENTION: New sequences of hepatitis C virus
; NUMBER OF SEQUENCES: 270
; STREET: Industriepark Zwijnaarde 7, box 4
; CITY: Ghent
; COUNTRY: Belgium
; ZIP: B-9052
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/638,693
; FILING DATE: 15-Aug-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE: 1995-01-11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 00 32 9 241 07 11
; TELEFAX: 00 32 9 241 07 99
; INFORMATION FOR SEQ ID NO: 158:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 158:
US-09-638-693-158

Query Match          45.0%; Score 9; DB 5; Length 176;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      7 HKFNSSGCP 15
Db      118 HKFNSSGCP 126
|||||

RESULT 15
US-09-873-224A-158
; Sequence 158, Application US/09873224A
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: New sequences of hepatitis C virus genotypes for diagnosis, prophylaxis and therapy
; FILE REFERENCE: 35
; CURRENT APPLICATION NUMBER: US/09/873,224A
; CURRENT FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 158
; LENGTH: 176
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-09-873-224A-158

Query Match          45.0%; Score 9; DB 5; Length 176;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      7 HKFNSSGCP 15
Db      118 HKFNSSGCP 126
|||||
```

Search completed: November 21, 2003, 22:12:54
Job time : 9.55 secs

0

6

0

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:46:27 ; Search time 9.25 Seconds
(without alignments)
207.932 Million cell updates/sec

Title: US-09-973-025-75

Perfect score: 20

Sequence: 1 AGLIYQHKFNSSGCPERLAS 20

Scoring table: OLIGO 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_76:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	3011	1	GNWVCH
2	16	80.0	3011	1	genome polyprotein
3	14	70.0	138	2	envelope protein -
4	14	70.0	513	2	structural protein
5	14	70.0	3011	1	GNWVC3
6	12	60.0	138	2	envelope protein -
7	12	60.0	3033	1	GNWVJ8
8	11	55.0	133	2	E48776
9	11	55.0	138	2	S24084
10	11	55.0	138	2	S24070
11	11	55.0	138	2	S24077
12	11	55.0	138	2	S24094
13	11	55.0	138	2	S24104
14	11	55.0	138	2	S24082
15	11	55.0	234	2	S23742
16	11	55.0	550	2	JH0711
17	11	55.0	716	2	JQ1366
18	11	55.0	3010	1	A45573
19	11	55.0	3014	1	JC5620
20	10	50.0	133	2	C48776
21	10	50.0	133	2	D48776
22	10	50.0	138	2	S24071
23	10	50.0	139	2	S24072
24	10	50.0	874	2	JQ0883
25	9	45.0	127	2	B48776
26	9	45.0	138	2	S24096
27	9	45.0	138	2	S24078
28	9	45.0	138	2	S24092
29	9	45.0	138	2	S24093

30	9	45.0	138	2	S24107	envelope protein -
31	9	45.0	138	2	S24106	envelope protein -
32	9	45.0	235	2	S32747	genome polyprotein
33	9	45.0	237	2	S32744	genome polyprotein
34	9	45.0	513	2	PC1284	genome polyprotein
35	9	45.0	874	2	JQ0881	genome polyprotein
36	9	45.0	876	2	PC2219	polypeptide - hepa
37	9	45.0	3010	1	GNWVTC	genome polyprotein
38	9	45.0	3033	1	JQ1303	genome polyprotein
39	8	40.0	138	2	PC1205	envelope protein -
40	8	40.0	138	2	PC1203	envelope protein -
41	8	40.0	138	2	PC1207	envelope protein -
42	8	40.0	138	2	PC1208	envelope protein -
43	8	40.0	138	2	PC1209	envelope protein -
44	8	40.0	138	2	PC1210	envelope protein -
45	8	40.0	138	2	PC1212	envelope protein -

ALIGNMENTS

RESULT 1

S40770

genome polyprotein - hepatitis C virus

N:Contains: capsid protein C; envelope protein M; hepatitisvirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001

C:Accession: S40770; PC1285

R:Okamoto, H.

submitted to the EMBL Data Library, March 1992

A:Reference number: S40770

A:Accession: S40770

A:Molecule type: genomic RNA

A:Residues: 1-3011 <OKA>

A:Cross-references: EMBL:D10749; NID:G221586; PIDN:BAA01582.1; PID:G221587

R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,

Jpn. J. Exp. Med. 60, 167-177, 1990

A:Title: The 5'-terminal sequence of the hepatitis C virus genome.

A:Reference number: PC1284; MUID:91013116; PMID:2170712

A:Accession: PC1285

A:Molecule type: genomic RNA

A:Residues: 1-513 <OK2>

A:Cross-references: GB:D00831; NID:G221511; PIDN:BAA00705.1; PID:G221512

A:Experimental source: isolate HC-J1

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin

F:116-191/Product: capsid protein C #status predicted <CPC>

F:192-389/Product: envelope protein M #status predicted <EPM>

F:390-729/Product: major envelope protein E #status predicted <MEE>

F:730-1006/Product: nonstructural protein NS1 #status predicted <NS1>

F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 100.0%; Score 20; DB 1; Length 3011;

Best Local Similarity 100.0%; Pred. No. 1.9e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGLIYQHKFNSSGCPERLAS 20

DB 439 AGLIYQHKFNSSGCPERLAS 458

RESULT 2

GNWVCH

genome polyprotein - hepatitis C virus (strain H)

N:Contains: capsid protein C; envelope protein M; hepatitisvirin (EC 3.4.21.98) (nonstructural

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C>Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A36814; A41546
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: c
A:Reference number: A36814
A:Accession: A36814
A:Molecule type: genomic RNA
A:Residues: 1-3011 <INC>
A:Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari
A:Reference number: A41546; MUID:92052256; PMID:1658800
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <BPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis B virus surface antigen (HBsAg) #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23
Query Match 80.0%; Score 16; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 4.3e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 YQKFNSSGCPERLAS 20
DB 443 YQKFNSSGCPERLAS 458
RESULT 3
S24081
envelope protein - hepatitis C virus (fragment)
N:Contains: envelope protein gp35; envelope protein gp70
C:Species: hepatitis C virus
C>Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C:Accession: S24081
R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraishi, K.; Ohkoshi,
Virus Res. 22, 107-123, 1992
A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C viru
A:Reference number: S24068; MUID:92230394; PMID:1314471
A:Accession: S24081
A:Molecule type: genomic RNA
A:Residues: 1-138 <NAT>
A:Cross-references: EMBL:X60569
A:Experimental source: isolate RE43
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; glycoprotein; polyprotein
F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:116-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted
F:49,55,80/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 70.0%; Score 14; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 HKFNSSGCPERLAS 20

Db 77 HKFNSSGCPERLAS 90
RESULT 4
A44150
structural protein - hepatitis C virus
C:Species: hepatitis C virus
C>Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 17-Nov-2000
C:Accession: A44150
R:Ching, W.M.; Wychowski, C.; Beach, M.J.; Wang, H.; Davies, C.L.; Carl, M.; Bradley, D.
Proc. Natl. Acad. Sci. U.S.A. 89, 3190-3194, 1992
A:Title: Interaction of immune sera with synthetic peptides corresponding to the structu
A:Reference number: A44150; MUID:92228749; PMID:1373489
A:Accession: A44150
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: genomic RNA
A:Residues: 1-513 <CHI>
C:Superfamily: hepatitis C virus genome polyprotein
Query Match 70.0%; Score 14; DB 2; Length 513;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 HKFNSSGCPERLAS 20
DB 445 HKFNSSGCPERLAS 458
RESULT 5
GNWVC3
genome polyprotein - hepatitis C virus (strain HCV-1)
N:Contains: capsid protein C; envelope protein M; hepatitis B virus surface antigen (HBsAg) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001
C:Accession: A39166; PQ0403; PQ0404
R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A:Title: Genetic organization and diversity of the hepatitis C virus.
A:Reference number: A39166; MUID:91172826; PMID:1848704
A:Accession: A39166
A:Molecule type: mRNA
A:Residues: 1-3011 <CHO>
A:Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L.
J. Gen. Virol. 73, 1131-1141, 1992
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A:Reference number: PQ0393; MUID:92268871; PMID:1316939
A:Accession: PQ0403
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CHA>
A:Cross-references: DDBJ:D10128
A:Experimental source: isolates E-b16
A:Accession: PQ0404
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CH2>
A:Experimental source: isolates E-b17
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <BPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis B virus surface antigen (HBsAg) #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 70.0%; Score 14; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 6.3e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 HKFNSSGCPERLAS 20
 |||||
 Db 445 HKFNSSGCPERLAS 458

RESULT 6
 S24080
 envelope protein - hepatitis C virus (fragment)
 N:Contains: envelope protein gp35; envelope protein gp70
 C:Species: hepatitis C virus
 C:Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
 C:Accession: S24080
 R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraishi, K.; Ohkoshi, Y.; Virus Res. 22, 107-123, 1992
 A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
 A:Reference number: S24068; MUID:92230394; PMID:1314471
 A:Accession: S24080
 A:Molecule type: genomic RNA
 A:Residues: 1-138 <KAT>
 A:Cross-references: EMBL:X60558
 A:Experimental source: isolate RE42
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: envelope protein; glycoprotein; polypeptide
 F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
 F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
 F:16-42/Region: hypervariable 1 #status predicted
 F:106-112/Region: hypervariable 2 #status predicted
 F:49,55,62,80/Binding site: carbohydrate (Asn) #status predicted

Query Match 60.0%; Score 12; DB 2; Length 138;
 Best Local Similarity 100.0%; Pred. No. 5.9e-06;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 FNSSGCPERLAS 20
 |||||
 Db 79 FNSSGCPERLAS 90

RESULT 7
 GNMVJ8
 genome polypeptide - hepatitis C virus (strain HC-J8)
 N:Contains: capsid protein C; envelope protein M; Hepacivirus (EC 3.4.21.98) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C:Accession: A40250; PQ0397; PQ0559
 R:Okamoto, H.; Kurai, K.; Okada, S.I.; Yanamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.; Virology 188, 331-341, 1992
 A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to reported sequences
 A:Reference number: A40250; MUID:92230232; PMID:1314459
 A:Accession: A40250
 A:Molecule type: genomic RNA
 A:Residues: 1-3033 <OKA>
 A:Cross-references: GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1; PID:g221609
 R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.I. J. Gen. Virol. 73, 1131-1141, 1992
 A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to other hepatitis C virus types
 A:Reference number: PQ0393; MUID:92268871; PMID:1316939
 A:Accession: PQ0397
 A:Molecule type: genomic RNA
 A:Residues: 2678-2754 <CHA>
 A:Cross-references: DBJ:D10134
 A:Experimental source: isolate B-12
 R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno, K.; Bloch, B. Biophys. Res. Commun. 181, 279-285, 1991
 A:Title: Distribution of plural HCV types in Japan.
 A:Reference number: PQ0554; MUID:92068204; PMID:1720309

A:Accession: PQ0559
 A:Molecule type: mRNA
 A:Residues: 2678-2729 <KAT>
 A:Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein
 F:116-191/Product: capsid protein C #status predicted <CPC>
 F:192-389/Product: envelope protein M #status predicted <BPM>
 F:390-733/Product: major envelope protein E #status predicted <MEE>
 F:734-1010/Product: nonstructural protein NS1 #status predicted <NS1>
 F:1011-1619/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1234-1241/Region: nucleotide-binding motif A (P-loop)
 F:1316-1321/Region: nucleotide-binding motif B
 F:1320-1323/Region: DEXH motif
 F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,233,299,305,417,423,430,448,477,534,542,558,576,627,649,1091,1217,1259,2038,22

Query Match 60.0%; Score 12; DB 1; Length 3033;
 Best Local Similarity 100.0%; Pred. No. 9.4e-05;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 HKFNSSGCPERL 18
 |||||
 Db 445 HKFNSSGCPERL 456

RESULT 8
 E48776
 polypeptide (E2/NS1 region, HVR1, HVR2) - hepatitis C virus (fragment)
 C:Species: hepatitis C virus
 C:Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 17-Nov-2000
 C:Accession: E48776
 R:Higashi, Y.; Kakumu, S.; Yoshioka, K.; Wakita, T.; Mizokami, M.; Ohba, K.; Ito, Y.; Virology 197, 659-668, 1993
 A:Title: Dynamics of genome change in the E2/NS1 region of hepatitis C virus in vivo.
 A:Reference number: A48776; MUID:94069940; PMID:8249288
 A:Accession: E48776
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-133 <HIG>
 A:Experimental source: subtype III, patient YG
 A:Note: sequence extracted from NCBI backbone (NCBI:140218)
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: polypeptide

Query Match 55.0%; Score 11; DB 2; Length 133;
 Best Local Similarity 100.0%; Pred. No. 7e-05;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 KFNSSGCPERL 18
 |||||
 Db 75 KFNSSGCPERL 85

RESULT 9
 S24084
 envelope protein - hepatitis C virus (fragment)
 N:Contains: envelope protein gp35; envelope protein gp70
 C:Species: hepatitis C virus
 C:Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
 C:Accession: S24084
 R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraishi, K.; Ohkoshi, Y.; Virus Res. 22, 107-123, 1992
 A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
 A:Reference number: S24068; MUID:92230394; PMID:1314471
 A:Accession: S24084
 A:Molecule type: genomic RNA
 A:Residues: 1-138 <KAT>
 A:Cross-references: EMBL:X60557
 A:Experimental source: isolate RE5

C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: envelope protein; glycoprotein; polyprotein
F;1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F;16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F;16-42/Region: hypervariable 1 #status predicted
F;106-112/Region: hypervariable 2 #status predicted
F;49,55,62,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.0%; Score 11; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 7.2e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 HKFNSSGCPER 17
Db 77 HKFNSSGCPER 87
|||||

RESULT 10
S24070
envelope protein - hepatitis C virus (fragment)
N;Contains: envelope protein gp35; envelope protein gp70
C;Species: hepatitis C virus
C;Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C;Accession: S24070
R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, K.; Ohkoshi,
Virus Res. 22, 107-123, 1992
A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A;Reference number: S24068; MUID:92230394; PMID:1314471
A;Accession: S24070
A;Molecule type: genomic RNA
A;Residues: 1-138 <KAT>
A;Cross-references: EMBL:X60554
A;Experimental source: isolate RE3
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: envelope protein; glycoprotein; polyprotein
F;1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F;16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F;16-42/Region: hypervariable 1 #status predicted
F;106-112/Region: hypervariable 2 #status predicted
F;49,55,62,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.0%; Score 11; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 7.2e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 HKFNSSGCPER 17
Db 77 HKFNSSGCPER 87
|||||

RESULT 11
S24077
envelope protein - hepatitis C virus (fragment)
N;Contains: envelope protein gp35; envelope protein gp70
C;Species: hepatitis C virus
C;Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C;Accession: S24077
R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, K.; Ohkoshi,
Virus Res. 22, 107-123, 1992
A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A;Reference number: S24068; MUID:92230394; PMID:1314471
A;Accession: S24077
A;Molecule type: genomic RNA
A;Residues: 1-138 <KAT>
A;Cross-references: EMBL:X60565
A;Experimental source: isolate RE39
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: envelope protein; glycoprotein; polyprotein
F;1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F;16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F;16-42/Region: hypervariable 1 #status predicted
F;106-112/Region: hypervariable 2 #status predicted
F;49,55,62,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.0%; Score 11; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 7.2e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 HKFNSSGCPER 17
Db 77 HKFNSSGCPER 87
|||||

RESULT 12
S24094
envelope protein - hepatitis C virus (fragment)
N;Contains: envelope protein gp35; envelope protein gp70
C;Species: hepatitis C virus
C;Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C;Accession: S24094
R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, K.; Ohkoshi,
Virus Res. 22, 107-123, 1992
A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A;Reference number: S24068; MUID:92230394; PMID:1314471
A;Accession: S24094
A;Molecule type: genomic RNA
A;Residues: 1-138 <KAT>
A;Cross-references: EMBL:X60579
A;Experimental source: isolate RE62
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: envelope protein; glycoprotein; polyprotein
F;1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F;16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F;16-42/Region: hypervariable 1 #status predicted
F;106-112/Region: hypervariable 2 #status predicted
F;49,55,62,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.0%; Score 11; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 7.2e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HKFNSSGCPER 17
Db 77 HKFNSSGCPER 87
|||||

RESULT 13
S24104
envelope protein - hepatitis C virus (fragment)
N;Contains: envelope protein gp35; envelope protein gp70
C;Species: hepatitis C virus
C;Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C;Accession: S24104
R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, K.; Ohkoshi,
Virus Res. 22, 107-123, 1992
A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A;Reference number: S24068; MUID:92230394; PMID:1314471
A;Accession: S24104
A;Molecule type: genomic RNA
A;Residues: 1-138 <KAT>
A;Cross-references: EMBL:X60587
A;Experimental source: isolate RE73B
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: envelope protein; glycoprotein; polyprotein
F;1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F;16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F;16-42/Region: hypervariable 1 #status predicted
F;106-112/Region: hypervariable 2 #status predicted
F;49,55,62,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.0%; Score 11; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 7.2e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 HKFNSSGCPER 17
Db 77 HKFNSSGCPER 87
|||||

Db 77 HKFNSSGCPER 87

RESULT 14

S24082
envelope protein - hepatitis C virus (fragment)
N:Contains: envelope protein gp35; envelope protein gp70
C:Species: hepatitis C virus
C:Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C:Accession: S24082
R:Kato, N.; Ootsuayama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; MuraIso, K.; Ohkoshi,
Virus Res. 22, 107-123, 1992
A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A:Reference number: S24068; MUID:92230394; PMID:1314471
A:Accession: S24082
A:Molecule type: genomic RNA
A:Residues: 1-138 <KAT>
A:Cross-references: EMBL:X60555
A:Experimental source: isolate RE4A
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; glycoprotein; polyprotein
F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted
F:49,55,62,80,110/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.0%; Score 11; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 7.2e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 HKFNSSGCPER 17

Db 77 HKFNSSGCPER 87

RESULT 15

S32742
genome polyprotein - hepatitis C virus (isolate CR-1) (fragment)
N:Contains: envelope protein E2
C:Species: hepatitis C virus
A:Variety: isolate CR-1
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S32742
R:Roggendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
submitted to the EMBL Data Library, February 1993
A:Description: Variability of the envelope regions of HCV in European isolates and its
A:Reference number: S32741
A:Accession: S32742
A:Molecule type: genomic RNA
A:Residues: 1-234 <ROG>
A:Cross-references: EMBL:X72979; NID:g296102; PIDN:CAA51484.1; PID:g296103
A:Experimental source: isolate CR-1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; polyprotein
F:1-234/Product: envelope protein E2 #status predicted <MAT>

Query Match 55.0%; Score 11; DB 2; Length 234;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 HKFNSSGCPER 17

Db 76 HKFNSSGCPER 86

Search completed: November 21, 2003, 21:11:30
Job time : 10.25 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:52 ; Search time 4.85 Seconds
(without alignments)
193.925 Million cell updates/sec

Title: US-09-973-025-75
Perfect score: 20
Sequence: 1 AGLIYQHKNSSGCPERLAS 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	80.0	3011	1 POLG_HCVH	P27958 h genome po
2	14	70.0	3011	1 POLG_HCV1	P26664 h genome po
3	12	60.0	3033	1 POLG_HCVJ8	P26661 h genome po
4	11	55.0	3010	1 POLG_HCVJ7	Q00269 h genome po
5	10	50.0	737	1 POLG_HCVJ7	P27961 hepatitis c
6	9	45.0	513	1 POLG_HCVJ2	P27959 hepatitis c
7	9	45.0	3010	1 POLG_HCVBK	P26663 h genome po
8	9	45.0	3033	1 POLG_HCVJ6	P26660 h genome po
9	7	35.0	737	1 POLG_HCVJ5	P27960 hepatitis c
10	6	30.0	105	1 HXB4_BRARE	P22574 brachydanio
11	6	30.0	149	1 VLYS_BPPRD	P13559 bacterioph
12	6	30.0	273	1 FPG_STRMU	P55045 streptococc
13	6	30.0	316	1 THER_BACTH	P00800 bacillus th
14	6	30.0	320	1 PHMX_HUMAN	Q96qsl homo sapien
15	6	30.0	334	1 MTN3_NEILA	P24582 neisseria l
16	6	30.0	361	1 UXAI_BACHD	Q9kx28 bacillus ha
17	6	30.0	508	1 HMCW_PIG	O02734 sus scrofa
18	6	30.0	520	1 TYX2_HUMAN	Q9y458 homo sapien
19	6	30.0	551	1 NFRS_BACST	P41133 bacillus st
20	6	30.0	644	1 YGM4_YEAST	P53129 saccharomyc
21	6	30.0	787	1 SMO_HUMAN	Q99835 homo sapien
22	6	30.0	793	1 SMO_MOUSE	P56726 mus musculu
23	6	30.0	793	1 SMO_RAT	P97698 rattus norv
24	6	30.0	794	1 SMO_CHICK	O42224 gallus gall
25	6	30.0	1162	1 LEPR_MOUSE	P48356 mus musculu
26	6	30.0	1165	1 LEPR_HUMAN	P48357 homo sapien
27	6	30.0	3010	1 POLG_HCVJ4	P26662 h genome po
28	6	30.0	3010	1 POLG_HCVJW	P29846 h genome po
29	5	25.0	52	1 IF2A_RABIT	P83268 oryctolagus
30	5	25.0	69	1 IF2A_PIG	P20460 sus scrofa
31	5	25.0	73	1 YG62_YEAST	P53340 saccharomyc
32	5	25.0	82	1 COAB_BPPFI	P03621 bacterioph
33	5	25.0	85	1 Y9KD_STRLI	P22399 streptomyce

34 5 25.0 92 1 FERN_AZOVI
35 5 25.0 101 1 RL21_SULSO
36 5 25.0 111 1 ESRI_SHEEP
37 5 25.0 111 1 TRKA_AZOC
38 5 25.0 113 1 SMPA_ECOLI
39 5 25.0 113 1 Y182_AERPE
40 5 25.0 114 1 RLA2_PAPAR
41 5 25.0 119 1 ACPS_STAAM
42 5 25.0 119 1 ACPS_STAAM
43 5 25.0 120 1 R18E_THEAC
44 5 25.0 146 1 NUOB_ERWCA
45 5 25.0 154 1 FLAV_BUCAP

P11054 azotobacter
Q9uxeo sulfolobus
P49885 ovis aries
Q04856 azorhizobiu
P23089 escherichia
Q9yfr7 aeropyrum p
P41099 parthenium
Q99si4 staphylococ
Q9zah6 staphylococ
Q9h106 thermoplasma
O85273 erwinia car
Q8k9n5 buchnera ap

ALIGNMENTS

RESULT 1

ID POLG_HCVH STANDARD; PRT; 3011 AA.
AC P27958.jo
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
DE Hepatitis C virus (isolate H) (HCV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658800;
RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M., Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C virus: comparison with American and Japanese isolates."; Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RL [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX MEDLINE=97331322; PubMed=9187654;
RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain."; Nat. Struct. Biol. 4:463-467(1997).
RL [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE=98154321; PubMed=9493270;
RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A., Murcko M.A., Lin C., Caron P.N.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding."; Structure 6:89-100(1998).
RL Structure 6:89-100(1998).
CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE ACTIVATION OF NS3.
CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position. Cys or Thr in Pi and Ser or Ala in P1.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1

CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M67463; AAA45534.1; --
 CC PIR; A36814; GNWVCH.
 DR PDB; 1HEI; 25-NOV-98.
 DR PDB; 1A1V; 16-FEB-99.
 DR PDB; 1A1R; 17-JUN-98.
 DR MEROPS; S29.001; --
 DR MEROPS; U39.001; --
 DR TRANSFAC; T04155; --
 DR TRANSFAC; T04155; -- DEAD.
 DR InterPro; IPR001410; HCV capsid.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR007095; RNA pol DS ps.
 DR InterPro; IPR007094; RNA pol Psvir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 KW INIT_MET 1 1
 FT CHAIN 1 191
 FT CHAIN 192 383
 FT CHAIN 384 746
 FT CHAIN 747 809
 FT CHAIN 810 1026
 FT CHAIN 1027 1657
 FT CHAIN 1658 1711
 FT CHAIN 1712 1972
 FT CHAIN 1973 2420
 FT CHAIN 2421 3011
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209

FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 476 476
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT STRAND 1224 1226
 FT TURN 1232 1233
 FT TURN 1236 1238
 FT TURN 1239 1246
 FT TURN 1247 1248
 FT STRAND 1251 1255
 FT HELIX 1258 1271
 FT TURN 1272 1272
 FT STRAND 1277 1280
 FT TURN 1281 1282
 FT STRAND 1283 1285
 FT STRAND 1291 1295
 FT HELIX 1296 1301
 FT TURN 1302 1303
 FT STRAND 1312 1316
 FT TURN 1317 1319
 FT HELIX 1323 1335
 FT TURN 1336 1340
 FT STRAND 1343 1347
 FT TURN 1352 1353
 FT TURN 1360 1361
 FT STRAND 1362 1366
 FT STRAND 1368 1368
 FT STRAND 1373 1375
 FT TURN 1376 1377
 FT STRAND 1378 1380
 FT HELIX 1382 1385
 FT STRAND 1389 1393
 FT HELIX 1397 1409
 FT TURN 1410 1411
 FT STRAND 1414 1417
 FT TURN 1419 1420
 FT STRAND 1432 1436
 FT TURN 1438 1439
 FT STRAND 1456 1463
 FT STRAND 1471 1478
 FT STRAND 1480 1480
 FT HELIX 1481 1488
 FT TURN 1489 1490
 FT STRAND 1497 1501
 FT STRAND 1507 1507
 FT STRAND 1511 1511
 FT HELIX 1514 1527
 FT HELIX 1532 1544
 FT STRAND 1550 1550
 FT HELIX 1555 1564
 FT HELIX 1570 1578
 FT TURN 1579 1580
 FT HELIX 1584 1597
 FT TURN 1598 1598
 FT HELIX 1606 1611
 FT TURN 1614 1618
 FT STRAND 1622 1623
 FT STRAND 1627 1627
 FT STRAND 1635 1636
 FT HELIX 1640 1652
 SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;

80.0%; Score 16; DB 1; Length 3011;

Query Match

Best Local Similarity 100.0%; Pred. No. 1.7e-09; Mismatches 0; Gaps 0; Indels 0; Gaps 0;

OY 5 YOHKFNSSGCPERLAS 20
DB 443 YOHKFNSSGCPERLAS 458

RESULT 2
POLG_HCV1 STANDARD; PRT; 3011 AA.
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate 1) (HCV)
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91172826; PubMed=148704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA Gallegos C., Coit D., Moughton M., Barr P.J., Weiner A.J.,
RA Bradley D.W., Kuo G., Houghton M.;
RL "Genetic organization and diversity of the hepatitis C virus.";
RT Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC [RNA] (N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M62321; AAA45676.1; -;
DR PIR; A39166; GNMVC3.
DR PDB; 1A1V; 16-FEB-99.
DR PDB; 1HEI; 25-NOV-98.
DR MEROPS; S29.001; -;
DR MEROPS; U39.001; -;
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RDRP.

DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase C; 1.
DR Pfam; PF00998; Viral_RDRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXdc; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT ACT_SITE 1083 1083 POTENTIAL.
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327197 MW; 65P8C9447FCESAF9 CRC64;

Query Match 70.0%; Score 14; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 HKFNSSGCPERLAS 20
DB 445 HKFNSSGCPERLAS 458

RESULT 3

POLG_HCVJ8
ID POLG_HCVJ8 STANDARD; PRT; 3033 AA.
AC P26661;
DT 01-AUG-1992 (Rel. 23, Created)

01-AUG-1992 (Rel. 23, Last sequence update)
 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polypeptide [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J8) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11115;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kura K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
 RA Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes";
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polypeptide, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D10988; BAA01761.1; -;
 DR PIR: A40250; GNMWJ8.
 DR HSP: P27958; IHEI.
 DR MEROPS: S29.001; -;
 DR MEROPS: U39.001; -;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RDRP.
 DR InterPro: IPR007095; RNA pol_PS.
 DR InterPro: IPR007094; RNA pol_psvir.
 DR Pfam: PF01543; HCV capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; Viral_RDRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.

DR SMART; SM00487; DEXdc; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2359 2359
 FT CARBOHYD 2811 2811
 FT CARBOHYD 3033 3033
 SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;
 Query Match 60.0%; Score 12; DB 1; Length 3033;
 Best Local Similarity 100.0%; Pred. No. 3.9e-05;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 7 HKFNSSGCPRL 18
 Db 445 HKFNSSGCPRL 456
 RESULT 4
 POLG HCVJT STANDARD; PRT; 3010 AA.
 AC Q00269;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polypeptide [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-JT) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92295714; PubMed=1318627;

RA Tanaka T., Kato N., Nakagawa M., Ootsubayama Y., Cho M.J.,
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 RT "Molecular cloning of hepatitis C virus genome from a single Japanese
 RT carrier: sequence variation within the same individual and among
 RT infected individuals.";
 RL Virus Res 23:39-53(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polypeptide, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {N} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MENA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, D11168; BAA01943.1; -;
 DR PIR; A45573; A45573.
 DR PDB; 1AIQ; 25-MAR-98.
 DR PDB; 1JXP; 14-JAN-98.
 DR MEROPS; S29.001; -;
 DR MEROPS; U39.001; -;
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RGRP.
 DR InterPro; IPR007095; RNA_pol_DS_Ps.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXdc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).

FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 POTENTIAL.
 FT ACT_SITE 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 ATP (POTENTIAL).
 FT NP_BIND 1316 DECH BOX.
 FT CARBOHYD 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;
 Query Match 55.0%; Score 11; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.00047;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 7 HKFNSGCPER 17
 Db 445 HKFNSGCPER 455
 RESULT 5
 POLG_HCVJ7 STANDARD; PRT; 737 AA.
 ID POLG_HCVJ7
 AC P27961;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
 DE protein (Envelope protein M); Major envelope protein E; Nonstructural
 DE proteins NS1 and NS2] (Fragment).
 OS Hepatitis C virus (isolate HC-J7) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11114;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishihiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MENA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

DE NS5B (p66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11105;
RX [1]
RP SEQUENCE FROM N.A.
RP MEDLINE=91140698; PubMed=1847440;
RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA Onishi E., Andoh T., Yoshida I., Okayama H.;
RA "Structure and organization of the hepatitis C virus genome isolated
RT from human carriers.";
RL J. Virol. 65:1105-1113 (1991).
RN [2]
RN SEQUENCE OF 1487-1500.
RP MEDLINE=96235224; PubMed=86471104;
RX Borowski P., Heiland M., Oehlmann K., Becker B., Kornetevy L.;
RA "Non-structural protein 3 of hepatitis C virus inhibits
RT phosphorylation mediated by cAMP-dependent protein kinase.";
RL Eur. J. Biochem. 237:611-618 (1996).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX MEDLINE=97015088; PubMed=8861916;
RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
RA Moomaw E.W., Adachi T., Hostomska Z.;
RA "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT trypsin-like fold and a structural zinc binding site.";
RL Cell 87:331-342 (1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX MEDLINE=98227846; PubMed=9568891;
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
RA "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
RL Protein Sci. 7:837-847 (1998).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA) (N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M58335; AA072945.1; -;
CC PIR; A38465; GNVVTC.
CC PDB; 1AIQ; 25-MAR-98.
CC PDB; 1JXP; 14-JAN-98.
CC PDB; 1NS3; 08-APR-98.
CC PDB; 1C2P; 15-NOV-00.
CC PDB; 1CSJ; 08-NOV-99.
CC PDB; 1GX5; 09-APR-02.
CC PDB; 1GX6; 10-APR-02.
CC PDB; 1QUV; 26-JUN-00.
CC PDB; 8OHM; 20-APR-99.
CC MEROPS; S29.001; -;
CC MEROPS; S29.001; -;
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002522; HCV_capsid.

DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXdc; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT ACT_SITE 1083 CHARGE RELAY SYSTEM.
FT ACT_SITE 1107 CHARGE RELAY SYSTEM.
FT ACT_SITE 1165 CHARGE RELAY SYSTEM.
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1031 1035
FT HELIX 1039 1047
FT STRAND 1050 1050
FT STRAND 1059 1063
FT STRAND 1068 1074
FT TURN 1075 1076
FT STRAND 1077 1081
FT HELIX 1082 1085
FT TURN 1086 1087

FT STRAND 1090 1092
 FT TURN 1093 1094
 FT STRAND 1095 1097
 FT STRAND 1101 1103
 FT TURN 1104 1107
 FT STRAND 1108 1112
 FT STRAND 1120 1120
 FT STRAND 1122 1122
 FT STRAND 1129 1133
 FT TURN 1135 1136
 FT STRAND 1139 1144
 FT STRAND 1149 1157
 FT HELIX 1158 1161
 FT TURN 1162 1163
 FT TURN 1165 1166
 FT STRAND 1168 1171
 FT TURN 1172 1174
 FT STRAND 1175 1186
 FT TURN 1187 1188
 FT STRAND 1189 1197
 FT HELIX 1198 1202
 FT TURN 1203 1204
 FT STRAND 1680 1688
 SQ SEQUENCE 3010 AA; 327189 MW; F84222D5ECFCDFD9C CRC64;

Query Match 45.0%; Score 9; DB 1; Length 3010;

Best Local Similarity 100.0%; Pred. No. 0.071; Gaps 0;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 FNSGCPER 17

Db 447 FNSGCPER 455

RESULT 8

ID POLG HCVJ6 STANDARD; PRT; 3033 AA.
 AC P26660;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J6) (HCV).
 OC Virus; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OK NCBI_TaxID=11113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92044440; PubMed=1658196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 RA Machida A., Miyakawa Y., Mayumi M.;
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 RT from a human carrier: comparison with reported isolates for conserved
 RT and divergent regions.";
 RL J. Gen. Virol. 72:2697-2704(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC [RNA] (N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.

CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

DR EMBL; D00944; BAA00792.1; -;
 DR PIR; JQ1303; JQ1303.
 DR HSSP; P27958; IHEI.
 DR MEROPS; Q529.001; -;
 DR MEROPS; U39.001; -;
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4A.
 DR InterPro; IPR001490; HCV NS4B.
 DR InterPro; IPR002868; HCV NS5A.
 DR InterPro; IPR002166; HCV RdRP.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01540; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4A; 1.
 DR Pfam; PF01001; HCV NS4B; 1.
 DR Pfam; PF01506; HCV NS5A; 1.
 DR Pfam; PF00271; helicase C; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_NET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT ACT_SITE 1234 1241
 FT NP_BIND 1320 1323
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542

FT	CHAIN	192	383	MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT	CHAIN	384	733	NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT	CHAIN	734	>737	NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT	TRANSMEM	347	369	POTENTIAL.
FT	CARBOHYD	196	196	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	209	209	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	234	234	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	305	305	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	417	417	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	423	423	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	430	430	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	448	448	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	477	477	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	534	534	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	542	542	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	558	558	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	578	578	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	627	627	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	649	649	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	NON TER	737	737	
FT	SEQUENCE	737 AA;	91207 MW; 3AF699D82AD501B1 CRC64;	
Query Match 35.0%; Score 7; DB 1; Length 737;				
Best Local Similarity 100.0%; Pred. No. 3.1;				
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	9 FNSSGCP 15			
	o			
DB	447 FNSSGCP 453			
RESULT 10				
ID	HXB4 BRARE STANDARD; PRT; 105 AA.			
AC	P22574; O42369;			
DT	01-AUG-1991 (Rel. 19, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Homeobox protein Hox-B4 (ZF-13) (Fragment).			
GN	HOXB4 OR HOXB4A OR HOX-B4 OR ZF-13.			
OS	Brachydanio rerio (Zebrafish) (Danio rerio).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;			
OC	Cyprinidae; Danio.			
OX	NCBI_TaxID=7955;			
OX	[1]			
RP	SEQUENCE OF 1-86 FROM N.A.			
RP	MEDLINE=89211958; PubMed=2468579;			
RA	Njelstad P.R., Molven A., Eiken H.G., Fjose A.;			
RT	"Structure and neural expression of a zebrafish homeobox sequence.";			
RL	Gene 73:33-46(1988).			
RN	[2]			
RP	SEQUENCE OF 44-105 FROM N.A.			
RA	Prince V.E., Moens C.B., Kimmel C.B., Ho R.K.;			
RT	"Zebrafish hox genes: expression in the hindbrain region of wild-type			
RT	and mutants of the segmentation gene, valentino.";			
RL	Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF			
CC	A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH			
CC	SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.			
CC	-1- SUBCELLULAR LOCATION: Nuclear.			
CC	-1- SIMILARITY: BELONGS TO THE ANTP HOMEBOX FAMILY.			
CC	"DEFORMED" SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to			

DR EMBL; Y13946; CAA74284.1; -;
DR PIR; J0489; J0489;
DR HSP; P02833; 9ANT;
DR ZFIN; ZDB-GENE-990415-105; hoxb4.
DR InterPro; IPR001827; Antennapedia.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD00010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
DR PROSITE; PS00032; ANTENNAPEDIA; PARTIAL.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT NON TER 1
FT DNA BIND 10 69
SQ SEQUENCE 105 AA; 12262 MW; BOEFD84D909289F1 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 105;
Best Local Similarity 100.0%; Pred. No. 6.8; Mismatches 0; Indels 0; Gaps 0;
Matches 6; Conservative 0;

QY 10 NSSGCP 15
DB 84 NSSGCP 89

RESULT 11
VLVS BPPRD STANDARD; PRT; 149 AA.
AC P13559;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Lytic enzyme (Protein p15).
GN XV.
OS Bacteriophage PRD1.
OC Viruses; dsDNA viruses, no RNA stage; Tectiviridae; Tectivirus.
OX NCBI_TaxID=10658;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89210844; PubMed=2651121;
RA Pakula T.M., Savilahti H., Bamford D.H.;
RT "Comparison of the amino acid sequence of the lytic enzyme from
broad-host-range bacteriophage PRD1 with sequences of other
cell-wall-peptidoglycan lytic enzymes.";
RL Eur. J. Biochem. 180:149-152(1989).
RN [2]
RP SEQUENCE OF 1-98 FROM N.A.
RX MEDLINE=88112855; PubMed=3322943;
RA Savilahti H., Bamford D.H.;
RT "The complete nucleotide sequence of the left very early region of
Escherichia coli bacteriophage PRD1 coding for the terminal protein
and the DNA polymerase.";
RL Gene 57:121-130(1987).
CC -!- FUNCTION: COMPLETE THE PRD1 LIFE CYCLE BY HOST CELL LYSIS.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; X14980; CAA33104.1; -;
DR EMBL; M69077; AAA32457.1; -;
DR EMBL; M22161; AAA32451.1; -;
DR FIR; S03568; LfBFD1.
KW Phage lysis protein.
SQ SEQUENCE 149 AA; 17269 MW; 50ABB718032939B5 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 9.3; Mismatches 0; Indels 0; Gaps 0;
Matches 6; Conservative 0;

QY 15 PERLAS 20
DB 112 PERLAS 117

RESULT 12
FPG STRMU STANDARD; PRT; 273 AA.
AC P55045;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Formamidopyrimidine-DNA glycosylase (EC 3.2.2.23) (FAPY-DNA
glycosylase).
GN MUTM OR FPG OR SMU.1614.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=GS-5;
RA Takamatsu N., Yamashita Y., Takehara T., Kuramitsu H.K.;
RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UA159 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -!- FUNCTION: THIS ENZYME MAY PLAY A SIGNIFICANT ROLE IN PROCESSES
LEADING TO RECOVERY FROM MUTAGENESIS AND/OR CELL DEATH BY
ALKYLATING AGENTS.
CC
CC -!- CATALYTIC ACTIVITY: Hydrolysis of DNA containing ring-opened N(7)-
methylguanine residues, releasing 2,6-diamino-4-hydroxy-5-(N-
methyl)formamidopyrimidine.
CC
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC
CC -!- SIMILARITY: BELONGS TO THE FPG FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; D26071; BAA05066.1; -;
DR EMBL; AE014992; AAN59255.1; -;
DR HSP; O50606; 1E8.
DR HAWAP; MF_00103; -; 1.
DR InterPro; IPR000191; Fapy_DNA_glyco.
DR InterPro; IPR000214; Fapy_DNAGlyco_zn.
DR Pfam; PF01149; Fapy_DNA_glyco; 1.
DR ProDom; PD003680; Fapy_DNA_glyco; 1.
DR TIGRFAMs; TIGR00577; fpg; 1.
DR PROSITE; PS01242; FPG; 1.
DR DNA repair; Hydrolase; Glycosidase; Zinc; Zinc-finger;
KW Complete proteome.
FT ZN.FING 247 270 POTENTIAL.
FT CONFLICT 195 195 K -> T (IN REF. 1).
SQ SEQUENCE 273 AA; 30875 MW; 4A9ECE1C00D9C484 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 273;
Best Local Similarity 100.0%; Pred. No. 16;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 PERLAS 20
|||||
Db 186 PERLAS 191

RESULT 13

1D THER BACTH STANDARD; PRT; 316 AA.

AC P00800;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Thermolysin (EC 3.4.24.27).

OS Bacillus thermoproteolyticus.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=1427;

RN [1]

RP SEQUENCE.

RA Titani K., Hermodson M.A., Ericsson L.H., Walsh K.A., Neurath H.;

RT "Amino-acid sequence of thermolysin.";

RL Nature New Biol. 238:35-37(1972).

RN [2]

RP ACTIVE SITE.

RX MEDLINE=74052951; PubMed=4808703;

RA Buxstein Y., Walsh K.A., Neurath H.;

RT "Evidence of an essential histidine residue in thermolysin.";

RL Biochemistry 13:205-210(1974).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).

RX MEDLINE=83085812; PubMed=7175940;

RA Holmes M.A., Matthews B.W.;

RT "Structure of thermolysin refined at 1.6-A resolution.";

RL J. Mol. Biol. 160:623-639(1982).

RN [4]

RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).

RX MEDLINE=75041142; PubMed=4214815;

RA Matthews B.W., Weaver L.H., Kester W.R.;

RT "The conformation of thermolysin.";

RL J. Biol. Chem. 249:8030-8044(1974).

RN [5]

RP STRUCTURE BY NMR OF 255-316.

RX MEDLINE=95086079; PubMed=7993910;

RA Rico M., Jimenez M.A., Gonzalez C., de Filippis V., Fontana A.;

RT "NMR solution structure of the C-terminal fragment 255-316 of

thermolysin: a dimer formed by subunits having the native

structure.";

RL Biochemistry 33:14834-14847(1994).

RN [6]

RP STRUCTURE BY NMR OF 205-316.

RX MEDLINE=97452621; PubMed=9305992;

RA Conejero-Lara F., Gonzalez C., Jimenez M.A., Padmanabhan S.,

RA Mateo P.L., Rico M.;

RT "NMR solution structure of the 205-316 C-terminal fragment of

thermolysin. An example of dimerization coupled to partial

unfolding.";

RL Biochemistry 36:11975-11983(1997).

CC -1- FUNCTION: THERMOSTABLE EXTRACELLULAR ZINC METALLOPROTEASE.

CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Xaa-[Leu > Xaa]-|Phe.

CC -1- COPACITOR: BINDS 1 ZINC ION AND FOUR CALCIUM IONS WHICH

CONTRIBUTE TO THE UNUSUAL THERMAL STABILITY OF THIS ENDOPEPTIDASE.

CC TWO ARE CLOSE TOGETHER AND PARTICIPATE IN A NETWORK OF SALT LINKS

ALSO INVOLVING ASP-138, GLU-177, LYS-182, ASP-185, GLU-190, AND

ASP-191. A THIRD INTERACTS WITH ASP-57 AND A FOURTH WITH ASP-200.

CC GLU AT POSITION 302 APPEARS TO FORM A SALT LINK WITH LYS-262.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M4.

DR PIR; 140579; HYBST.

DR PDB; 4TLN; 31-JAN-84.

DR PDB; 7TLN; 31-JAN-84.

DR PDB; 7TLN; 31-JAN-84.

DR PDB; 8TLN; 30-APR-94.

DR PDB; 1TMN; 09-JAN-89.
DR PDB; 2TMN; 15-JAN-91.
DR PDB; 3TMN; 09-JAN-89.
DR PDB; 4TMN; 09-JAN-89.
DR PDB; 5TMN; 09-JAN-89.
DR PDB; 6TMN; 09-JAN-89.
DR PDB; 1THL; 31-JAN-94.
DR PDB; 1TLP; 09-JAN-89.
DR PDB; 1HYT; 31-JUL-94.
DR PDB; 1TRL; 07-FEB-95.
DR PDB; 1LNA; 08-MAY-95.
DR PDB; 1LNB; 08-MAY-95.
DR PDB; 1LNC; 08-MAY-95.
DR PDB; 1LND; 08-MAY-95.
DR PDB; 1LNE; 08-MAY-95.
DR PDB; 1LNF; 08-MAY-95.
DR PDB; 1LFJ; 18-APR-01.
DR PDB; 1FJO; 18-APR-01.
DR PDB; 1FJU; 18-APR-01.
DR PDB; 1FJV; 18-APR-01.
DR PDB; 1FJW; 18-APR-01.
DR PDB; 1GXW; 05-DEC-02.
DR PDB; 1KEI; 16-NOV-02.
DR PDB; 1KJO; 05-DEC-02.
DR PDB; 1KJP; 05-DEC-02.
DR PDB; 1KKK; 10-DEC-02.
DR PDB; 1KL6; 11-DEC-02.
DR PDB; 1KR6; 09-JAN-03.
DR PDB; 1KRO; 10-JAN-03.
DR PDB; 1KST; 11-JAN-03.
DR PDB; 1KTO; 17-JAN-03.
DR PDB; 1L3F; 03-JUL-02.
DR PDB; 1OSO; 25-MAR-03.
DR PDB; 1QF0; 29-DEC-99.
DR PDB; 1QF1; 29-DEC-99.
DR PDB; 1QF2; 20-DEC-00.
DR PDB; 1TKI; 18-FEB-00.
DR PDB; 2TLI; 13-MAR-00.
DR PDB; 3TLI; 13-MAR-00.
DR PDB; 4TLI; 13-MAR-00.
DR PDB; 5TLI; 13-MAR-00.
DR PDB; 6TLI; 13-MAR-00.
DR PDB; 7TLI; 13-MAR-00.
DR PDB; 8TLI; 13-MAR-00.
DR MEROPS; M04.001; -.
DR InterPro; IPR001570; Peptidase M4.
DR InterPro; IPR006025; Zn MTpeptdse.
DR Pfam; PF01447; Peptidase M4; 1.
DR Pfam; PF02868; Peptidase M4_C; 1.
DR PRINTS; PR00730; THERMOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Calcium; 3D-structure.
FT METAL 142 142 ZINC (CATALYTIC).
FT ACT SITE 143 143 ZINC (CATALYTIC).
FT METAL 146 146 ZINC (CATALYTIC).
FT METAL 166 166 ZINC (CATALYTIC).
FT ACT SITE 231 231 PROTON DONOR.
FT STRAND 4 11
FT TURN 13 14
FT STRAND 17 24
FT STRAND 28 29
FT STRAND 31 32
FT TURN 33 34
FT TURN 36 37
FT STRAND 39 43
FT TURN 45 46
FT STRAND 53 54
FT STRAND 56 57
FT STRAND 61 62
FT HELIX 65 67
FT HELIX 68 88

```

FT TURN 92 94
FT STRAND 100 104
FT STRAND 113 115
FT STRAND 120 122
FT STRAND 130 130
FT STRAND 133 135
FT HELIX 137 151
FT HELIX 152 153
FT TURN 159 180
FT HELIX 181 181
FT TURN 187 188
FT STRAND 190 192
FT STRAND 193 193
FT STRAND 195 196
FT TURN 198 199
FT STRAND 203 204
FT HELIX 208 211
FT TURN 212 212
FT HELIX 217 219
FT HELIX 225 229
FT TURN 230 233
FT HELIX 234 246
FT STRAND 248 250
FT TURN 251 252
FT STRAND 253 255
FT HELIX 260 273
FT TURN 274 274
FT TURN 277 278
FT HELIX 281 296
FT TURN 298 299
FT HELIX 301 312
FT TURN 313 314
SQ SEQUENCE 316 AA; 34333 MW; 004EDAF478744BB4 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 316;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGLIYQ 6
DB 153 AGLIYQ 158

RESULT 14
PHMX HUMAN STANDARD; PRT: 320 AA.
AC Q96QS1; Q96KX4; Q9HC50; Q9HC51; Q9Y5U1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Phenx protein.
GN PHMX OR TSSC6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A. (ISOFORM 2), AND TISSUE SPECIFICITY.
RP MEDLINE=99172079; PubMed=10072438;
RA Lee M.P., Brandenburg S., Landes G.M., Adams M., Miller G.,
RA Feinberg A.P.
RA "Two novel genes in the center of the 11p15 imprinted domain escape
RT genomic imprinting.";
RL Hum. Mol. Genet. 8:683-690(1999).
[2]
RN SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=21575689; PubMed=11718897;
RA Robb L., Tarrant J., Groom J., Ibrahim M., Li R., Borobakas B.,
RA Wright M.D.;
RA "Molecular characterisation of mouse and human TSSC6: evidence that
RT TSSC6 is a genuine member of the tetraspanin superfamily and is
RL expressed specifically in haematopoietic organs.";
RL Biochim. Biophys. Acta 1522:31-41(2001).

[3]
RN SEQUENCE FROM N.A. (ISOFORM 5).
RP TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skaleka U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[4]
RN SEQUENCE OF 33-320 FROM N.A. (ISOFORMS 3 AND 4), AND DEVELOPMENTAL
RP STAGE.
RX MEDLINE=20408882; PubMed=10950922;
RA Nicholson R.H., Pantano S., Eliason J.F., Galy A., Weiler S.,
RA Kaplan J., Hughes M.R., Ko M.S.H.;
RA "Phemx, a novel mouse gene expressed in hematopoietic cells maps to
RT the imprinted cluster on distal chromosome 7.";
Genomics 68:13-21(2000).
RL
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=5;
CC Comment=Additional isoforms seem to exist. Experimental
CC confirmation may be lacking for some isoforms;
CC Name=1;
CC IsoId=Q96QS1-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q96QS1-2; Sequence=VSP_003932;
CC Name=3;
CC IsoId=Q96QS1-3; Sequence=VSP_003937, VSP_003938;
CC Name=4;
CC IsoId=Q96QS1-4; Sequence=VSP_003932, VSP_003933, VSP_003934;
CC Name=5;
CC IsoId=Q96QS1-5; Sequence=VSP_003935, VSP_003936;
CC -!- TISSUE SPECIFICITY: Expressed ubiquitously at low levels. High
CC levels of expression are confined to hematopoietic tissues
CC including peripheral blood leukocytes, thymus and spleen.
CC -!- DEVELOPMENTAL STAGE: Expressed from early embryogenesis through to
CC adulthood.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF125569; AAD23580.1; -
CC EMBL; AY039001; AAK84431.1; -
CC EMBL; BC016693; AAH16693.1; -
CC EMBL; AF176070; AAG15840.1; -
CC EMBL; AF176071; AAG15841.1; -
CC Genew; HGNC:13410; PHMX.
CC MIM; 603853; -
CC GO; GO:0016021; C:integral to membrane; NAS.
CC GO; GO:0007267; P:cell-cell signaling; TAS.
CC Transmembrane; Alternative splicing.
CC TRANSMEM 14 34
CC POTENTIAL.
CC FT
```

```
FT TRANSMEM 60 80 POTENTIAL.
FT TRANSMEM 90 110 POTENTIAL.
FT TRANSMEM 203 223 POTENTIAL.
FT VARSPPLIC 1 30 Missing (in isoform 2 and isoform 4).
FT VARSPPLIC 137 154 /FTId=VSP 003932.
FT VARSPPLIC 154 154 GTSHVRRQELAAIQDVL -> VSVLWEEVSFQPSGEHRG
FT VARSPPLIC 154 154 (in isoform 4).
FT VARSPPLIC 155 320 /FTId=VSP 003933.
FT VARSPPLIC 155 320 Missing (in isoform 4).
FT VARSPPLIC 210 258 /FTId=VSP 003934.
FT VARSPPLIC 210 258 VSALLFSFLPAINRCGSLDRKGYTLTPRACGRQPOEPS
FT VARSPPLIC 259 320 LURCSQGG -> LGPQGHDPDPTSMWPPAPCAQPLEMLPG
FT VARSPPLIC 259 320 WTHLSPLRSCYWSKRMGLG (in isoform 5).
FT VARSPPLIC 241 253 /FTId=VSP 003935.
FT VARSPPLIC 241 253 Missing (in isoform 5).
FT VARSPPLIC 254 320 /FTId=VSP 003936.
FT VARSPPLIC 254 320 ACGRPQPSLLR -> SPGQKSRWAQWVP (in
FT VARSPPLIC 254 320 isoform 3).
FT VARSPPLIC 254 320 /FTId=VSP 003937.
FT VARSPPLIC 254 320 Missing (in isoform 3).
FT VARSPPLIC 254 320 /FTId=VSP 003938.
SQ SEQUENCE 320 AA; 34631 MW; 534E4C29767B8E6D CRC64;

Query Match 30.0%; Score 6; DB 1; Length 320;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 SGCPR 17
DB 311 SGCPR 316

RESULT 15
MTN3 NEILA
ID MTN3 NEILA STANDARD; PRT; 334 AA.
AC P24582;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Modification methylase NlaIII (EC 2.1.1.72) (Adenine-specific
DE methyltransferase NlaIII) (M.NlaIII).
GN NLAIII.
OS Neisseria lactamica.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=486;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 23970 / NRCC 2118;
RX MEDLINE=91117164; PubMed=2277628;
RA Labbe D., Hoeltke H.J., Lau P.C.K.;
RT "Cloning and characterization of two tandemly arranged DNA
RT methyltransferase genes of Neisseria lactamica: an adenine-specific
RT M.NlaIII and a cytosine-type methylase.";
RL Mol. Gen. Genet. 224:101-110(1990).
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC CATG, CAUSES SPECIFIC METHYLATION ON A-2 ON BOTH STRANDS AND
CC PROTECTS THE DNA FROM CLEAVAGE BY THE NlaIII ENDONUCLEASE.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-57 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54485; CAA38356.1; -
DR PIR; S12036; XYNHAL.
DR REBASE; 3468; M.NlaIII.
```

```
DR InterPro; IPR002294; D12N6_mtfase.
DR InterPro; IPR002052; N6_Mtase.
DR Pfam; PF02086; Methyltransferase_D12; 1.
DR PRINTS; PR00505; D12N6_MTFRASE.
DR PROSITE; PS00092; N6_MTASE; 1.
KW Transferase; Methyltransferase; Restriction system.
SQ SEQUENCE 334 AA; 38382 MW; 8BCF1A1A35F02E89 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 334;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLIYQH 7
DB 93 GLIYQH 98

Search completed: November 21, 2003, 20:59:53
Job time : 4.85 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:45:57 ; Search time 24.4 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-75

Perfect score: 20

Sequence: 1 AGLIYQHKFNSSGCPERLAS 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- SPTREMBL_23.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phase.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	20	100.0	135	12 Q68196	Q68196 hepatitis c
2	20	100.0	513	12 Q81220	Q81220 hepatitis c
3	20	100.0	3011	12 Q03463	Q03463 hepatitis c
4	16	80.0	96	12 Q912N5	Q912N5 hepatitis c
5	16	80.0	130	12 Q912M6	Q912M6 hepatitis c
6	16	80.0	130	12 Q912P3	Q912P3 hepatitis c
7	16	80.0	130	12 Q912M5	Q912M5 hepatitis c
8	16	80.0	130	12 Q912N1	Q912N1 hepatitis c
9	16	80.0	130	12 Q912M7	Q912M7 hepatitis c
10	16	80.0	130	12 Q912N2	Q912N2 hepatitis c
11	16	80.0	130	12 Q912P5	Q912P5 hepatitis c
12	16	80.0	130	12 Q912Q4	Q912Q4 hepatitis c
13	16	80.0	130	12 Q912N6	Q912N6 hepatitis c
14	16	80.0	130	12 Q912M8	Q912M8 hepatitis c
15	16	80.0	130	12 Q912N4	Q912N4 hepatitis c
16	16	80.0	130	12 Q912M2	Q912M2 hepatitis c

17	16	80.0	130	12 Q912P0	Q912P0 hepatitis c
18	16	80.0	130	12 Q912P2	Q912P2 hepatitis c
19	16	80.0	130	12 Q912Q5	Q912Q5 hepatitis c
20	16	80.0	130	12 Q912N7	Q912N7 hepatitis c
21	16	80.0	130	12 Q912Q2	Q912Q2 hepatitis c
22	16	80.0	130	12 Q912N3	Q912N3 hepatitis c
23	16	80.0	130	12 Q912Q6	Q912Q6 hepatitis c
24	16	80.0	130	12 Q912M3	Q912M3 hepatitis c
25	16	80.0	130	12 Q912M1	Q912M1 hepatitis c
26	16	80.0	130	12 Q912Q3	Q912Q3 hepatitis c
27	16	80.0	130	12 Q912P1	Q912P1 hepatitis c
28	16	80.0	130	12 Q912N8	Q912N8 hepatitis c
29	16	80.0	130	12 Q912N0	Q912N0 hepatitis c
30	16	80.0	130	12 Q912M0	Q912M0 hepatitis c
31	16	80.0	130	12 Q912M4	Q912M4 hepatitis c
32	16	80.0	130	12 Q912P4	Q912P4 hepatitis c
33	16	80.0	130	12 Q912N9	Q912N9 hepatitis c
34	16	80.0	130	12 Q912M9	Q912M9 hepatitis c
35	16	80.0	186	12 Q912YH2	Q912YH2 hepatitis c
36	16	80.0	186	12 Q912YH2	Q912YH2 hepatitis c
37	16	80.0	186	12 Q912YH2	Q912YH2 hepatitis c
38	16	80.0	186	12 Q912YH2	Q912YH2 hepatitis c
39	16	80.0	186	12 Q912YH2	Q912YH2 hepatitis c
40	16	80.0	186	12 Q912YH2	Q912YH2 hepatitis c
41	16	80.0	186	12 Q912YH2	Q912YH2 hepatitis c
42	16	80.0	186	12 Q912YH2	Q912YH2 hepatitis c
43	16	80.0	186	12 Q912YH2	Q912YH2 hepatitis c
44	16	80.0	186	12 Q912YH2	Q912YH2 hepatitis c
45	16	80.0	186	12 Q912YH2	Q912YH2 hepatitis c

ALIGNMENTS

RESULT 1

Q68196 ID Q68196 PRELIMINARY; PRT; 135 AA.

AC Q68196;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Envelope protein (Genome polyprotein) (Fragment).
GN E2/NS1.

OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;

RP [1]

RV SEQUENCE FROM N.A.

RC STRAIN=1a;

RX MEDLINE=95146953; PubMed=7844535;

RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;

RT "Prevalence of hepatitis C virus sequence variants in South-East Asia.";

RL J. Gen. Virol. 76:211-215(1995).

DR EMBL; U14232; AAC53921.1; -.

DR InterPro; IPR002519; HCV env.

DR InterPro; IPR002531; HCV_NSI.

DR Pfam; PF01539; HCV env; 1.

DR Pfam; PF01560; HCV_NSI; 1.

DR ProDom; PD186062; HCV_NSI; 1.

KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;

Polyprotein; Transmembrane.

FT NON_TER 1

FT NON_TER 135 135

SQ SEQUENCE 135 AA; 14408 MW; AD9A1939B546B8D1 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 135;

Best Local Similarity 100.0%; Pred. No. 4.1e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLIYQHKFNSSGCPERLAS 20

|||||

```

Db          90 AGLIYQHKFNSSGCCPERLAS 109

RESULT 2
Q81220      PRELIMINARY;          PRT;      513 AA.
AC
Q81220;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Structural protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; asRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=91013116; PubMed=2170712;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA Yoshizawa H.;
RT "The 5'-terminal sequence of the hepatitis C virus genome.";
RL Jpn. J. Exp. Med. 60:167-177(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=92044440; PubMed=1658196;
RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
RA Miyakawa Y., Mayumi M.;
RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=94270990; PubMed=7545932;
RA Hotta H., Doi H., Hayashi T., Furwanta M., Soemarto W., Mizokami M.,
RA Ohba K., Homma M.;
RT "Analysis of the core and E1 envelope region sequences of a novel
RT variant of hepatitis C virus obtained in Indonesia.";
RL Arch. Virol. 136:53-62(1994).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; D00831; BAA00705.1; -.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 513
SQ SEQUENCE 513 AA; 55710 MW; B1B54EBF9C359407 CRC64;

Query Match      100.0%; Score 20; DB 12; Length 513;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGLIYQHKFNSSGCCPERLAS 20
    |||||
Db          439 AGLIYQHKFNSSGCCPERLAS 458

RESULT 3
Q03463
ID Q03463      PRELIMINARY;          PRT;      3011 AA.

```

```

AC Q03463;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; asRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=91013116; PubMed=2170712;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA Yoshizawa H.;
RT "The 5'-terminal sequence of the hepatitis C virus genome.";
RL Jpn. J. Exp. Med. 60:167-177(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=92044440; PubMed=1658196;
RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
RA Miyakawa Y., Mayumi M.;
RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=93117120; PubMed=1335573;
RA Okamoto H., Kanai N., Mishiro S.;
RT "Full-length nucleotide sequence of a Japanese hepatitis C virus
RT isolate (HC-J1) with high homology to USA isolates.";
RL Nucleic Acids Res. 20:6410-6410(1992).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RA Okamoto H.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=94174722; PubMed=7510436;
RA Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,
RA Inchausti G.;
RT "Characterization and mapping of a B-cell immunogenic domain in
RT hepatitis C virus E2 glycoprotein using a yeast peptide library.";
RL Virology 200:246-255(1994).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; D10749; BAA01582.1; -.
DR HSSP; P27958; 1HE1.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA_pol_Ds_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.

```

DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; Helicase C; 1.
 DR Pfam; PF00998; Viral RRP; 1.
 DR Pfam; PF0186062; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXdc; 1.
 DR PROSITE; PS50507; RDRP POSITIVE; 1.
 DR PROSITE; PS50521; RDRP VIRAL; 1.
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydroxylase; Nonstructural protein; Polyprotein;
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3011 AA; 327112 MW; 97E9052C0250463B CRC64;

Query Match 100.0%; Score 20; DB 12; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 7.2e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLYQHKFNSSGCPERLAS 20
 |||||
 Db 439 AGLYQHKFNSSGCPERLAS 458

RESULT 4

Q91ZNS PRELIMINARY; PRT; 96 AA.
 AC Q91ZNS;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=X186A-b;
 RX MEDLINE=20173694; PubMed=10708420;
 RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
 RA Wang Y.M., Thomas D.L.;
 RT "Hypervariable region 1 sequence stability during hepatitis C virus
 RT replication in chimpanzees";
 RL J. Virol. 74:3058-3066(2000).
 DR EMBL; AF230437; AAF68705.1; -.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 SQ SEQUENCE 96 AA; 10158 MW; 72211BCD78BDB768 CRC64;

Query Match 80.0%; Score 16; DB 12; Length 96;
 Best Local Similarity 100.0%; Pred. No. 6.6e-10;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 YQHKFNSSGCPERLAS 20
 |||||
 Db 71 YQHKFNSSGCPERLAS 86

RESULT 5

Q91ZM6 PRELIMINARY; PRT; 130 AA.
 AC Q91ZM6;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=X187A-e;
 RX MEDLINE=20173694; PubMed=10708420;
 RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
 RA Wang Y.M., Thomas D.L.;
 RT "Hypervariable region 1 sequence stability during hepatitis C virus
 RT replication in chimpanzees";
 RL J. Virol. 74:3058-3066(2000).
 DR EMBL; AF230446; AAF68714.1; -.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 SQ SEQUENCE 130 AA; 13899 MW; 7A833EEE13183D24 CRC64;

Query Match 80.0%; Score 16; DB 12; Length 130;
 Best Local Similarity 100.0%; Pred. No. 8.7e-10;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 YQHKFNSSGCPERLAS 20
 |||||
 Db 71 YQHKFNSSGCPERLAS 86

RESULT 6

Q91ZP3 PRELIMINARY; PRT; 130 AA.
 ID Q91ZP3;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=X268A-ff;
 RX MEDLINE=20173694; PubMed=10708420;
 RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
 RA Wang Y.M., Thomas D.L.;
 RT "Hypervariable region 1 sequence stability during hepatitis C virus
 RT replication in chimpanzees";
 RL J. Virol. 74:3058-3066(2000).
 DR EMBL; AF230429; AAF68697.1; -.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 SQ SEQUENCE 130 AA; 13905 MW; 963BCFAA4A304356E CRC64;

Query Match 80.0%; Score 16; DB 12; Length 130;
 Best Local Similarity 100.0%; Pred. No. 8.7e-10;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 YQHKFNSSGCPERLAS 20
 |||||
 Db 71 YQHKFNSSGCPERLAS 86

RESULT 7

Q91ZM5

```

ID Q91ZM5 PRELIMINARY; PRT; 130 AA.
AC Q91ZM5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X361A-a;
RX Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RA Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL; AF230447; AAF68715.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 130
SQ SEQUENCE 130 AA; 13915 MW; 6AD37BEE13042124 CRC64;

Query Match 80.0%; Score 16; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 8.7e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 YQHKFNSSGCGPERLAS 20
Db 71 YQHKFNSSGCGPERLAS 86

RESULT 8
Q91ZN1 PRELIMINARY; PRT; 130 AA.
AC Q91ZN1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X187HC4.1;
RX Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RA Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL; AF230441; AAF68709.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 130
SQ SEQUENCE 130 AA; 13875 MW; 663BDEA4A3043573 CRC64;

Query Match 80.0%; Score 16; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 8.7e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 5 YQHKFNSSGCGPERLAS 20
Db 71 YQHKFNSSGCGPERLAS 86

RESULT 9
Q91ZM7 PRELIMINARY; PRT; 130 AA.
AC Q91ZM7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X187A-ss;
RX Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RA Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL; AF230445; AAF68713.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 130
SQ SEQUENCE 130 AA; 13916 MW; 3A79744413094129 CRC64;

Query Match 80.0%; Score 16; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 8.7e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 YQHKFNSSGCGPERLAS 20
Db 71 YQHKFNSSGCGPERLAS 86

RESULT 10
Q91ZN2 PRELIMINARY; PRT; 130 AA.
AC Q91ZN2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X187HC3.11;
RX Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RA Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL; AF230440; AAF68708.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 130
SQ SEQUENCE 130 AA; 13875 MW; 663BDEA4A3043573 CRC64;

Query Match 80.0%; Score 16; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 8.7e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

FT NON_TER 130 130
SQ SEQUENCE 130 AA; 13915 MW; 6AD37EEE13042124 CRC64;

Query Match
Best Local Similarity 80.0%; Score 16; DB 12; Length 130;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 YQHKFNSSGCPERLAS 20
Db 71 YQHKFNSSGCPERLAS 86

RESULT 11
Q912P5 Q912P5 PRELIMINARY; PRT; 130 AA.
AC Q912P5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X268A-a;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RA Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
RT replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL; AF230427; AAF68695.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProbDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1 1
FT NON_TER 130 130
SQ SEQUENCE 130 AA; 13915 MW; 6AD37EEE13042124 CRC64;

Query Match
Best Local Similarity 80.0%; Score 16; DB 12; Length 130;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 YQHKFNSSGCPERLAS 20
Db 71 YQHKFNSSGCPERLAS 86

RESULT 12
Q912Q4 Q912Q4 PRELIMINARY; PRT; 130 AA.
AC Q912Q4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X007A-f;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RA Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
RT replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).

```

```

DR EMBL; AF230418; AAF68686.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProbDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1 1
FT NON_TER 130 130
SQ SEQUENCE 130 AA; 13869 MW; 6PD73EF4D302E7E4 CRC64;

Query Match
Best Local Similarity 80.0%; Score 16; DB 12; Length 130;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 YQHKFNSSGCPERLAS 20
Db 71 YQHKFNSSGCPERLAS 86

RESULT 13
Q912N6 Q912N6 PRELIMINARY; PRT; 130 AA.
AC Q912N6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X186A-11;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RA Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
RT replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL; AF230436; AAF68704.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProbDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1 1
FT NON_TER 130 130
SQ SEQUENCE 130 AA; 13915 MW; 6AD37EEE13042124 CRC64;

Query Match
Best Local Similarity 80.0%; Score 16; DB 12; Length 130;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 YQHKFNSSGCPERLAS 20
Db 71 YQHKFNSSGCPERLAS 86

RESULT 14
Q912M8 Q912M8 PRELIMINARY; PRT; 130 AA.
AC Q912M8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN=X187A-nn;
RX MEDLINE=20173694; PubMed=10708420;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RW Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL; AF230444; AAF68712.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 130 130
SQ SEQUENCE 130 AA; 13931 MW; 6AC73ABE13043A3F CRC64;

Query Match 80.0%; Score 16; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 8.7e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 YQHKFNSSGCPERLAS 20
Db 71 YQHKFNSSGCPERLAS 86
|||||

RESULT 15
Q91ZN4 PRELIMINARY; PRT; 130 AA.
AC Q91ZN4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_taxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X186A-gg;
RX MEDLINE=20173694; PubMed=10708420;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RW Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL; AF230438; AAF68706.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 130 130
SQ SEQUENCE 130 AA; 13857 MW; 6B037DEE13042224 CRC64;

Query Match 80.0%; Score 16; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 8.7e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 YQHKFNSSGCPERLAS 20
Db 71 YQHKFNSSGCPERLAS 86
|||||

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:22 ; Search time 32.1 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-76

Perfect score: 20

Sequence: 1 GCPERLASCRELTDFDQMG 20

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	16 AAR84505	Hepatitis C virus
2	20	100.0	20	17 AAR90999	HCV E2 peptide E2-
3	20	100.0	20	23 AAO18705	Hepatitis C virus
4	20	100.0	37	12 AAR13359	P447 HCV antigen (
5	20	100.0	37	14 AAR33878	Polypeptide p447 c
6	20	100.0	76	21 AAB18528	Protein encoded by
7	20	100.0	254	22 AAB68043	Amino acid sequenc
8	20	100.0	278	14 AAR33997	Th E2/NS1 protein.
9	20	100.0	363	22 AAB68042	Amino acid sequenc

10	20	100.0	402	14 AAR34438	Sequence of glycop
11	20	100.0	454	10 AAP90183	Sequence of hepati
12	20	100.0	454	10 AAP92049	Sequence encoded b
13	20	100.0	454	21 AAB18526	Protein encoded by
14	20	100.0	463	14 AAR33588	HCV CKS-NS181 fusi
15	20	100.0	463	14 AAR33187	Sequence of subfra
16	20	100.0	463	22 AAB69007	HCV recombinant an
17	20	100.0	480	14 AAR33992	HCV-1 E2/NS1 prote
18	20	100.0	531	22 AAE02622	Chimeric HCV E2661
19	20	100.0	621	14 AAR33185	Sequence of subfra
20	20	100.0	622	14 AAR33591	HCV CKS-NS181-NS1S
21	20	100.0	622	22 AAB69010	HCV recombinant an
22	20	100.0	637	24 ABP57410	Hepatitis C virus
23	20	100.0	663	17 AAR92935	HCV1 E2 + NS2 poly
24	20	100.0	663	20 AAW67615	Hepatitis C virus
25	20	100.0	738	14 AAR33592	HCV CKS-full lengt
26	20	100.0	738	22 AAB69011	HCV recombinant an
27	20	100.0	2435	13 AAR25135	HCV polypeptide 1.
28	20	100.0	2436	10 AAP92050	Sequence encoded i
29	20	100.0	2436	10 AAP90288	Peptide encoded by
30	20	100.0	2436	13 AAR28582	HCV amino acid seq
31	20	100.0	2772	11 AAR08123	Hepatitis C virus
32	20	100.0	2772	21 AAB18540	Protein encoded by
33	20	100.0	2816	14 AAR34009	HCV-1 polypeptin.
34	20	100.0	2894	13 AAR24440	Composite HCV HC-J
35	20	100.0	2894	16 AAR70230	Composite hepatiti
36	20	100.0	2955	11 AAR08124	Hepatitis C virus
37	20	100.0	2955	20 AAY14975	Amino acid sequenc
38	20	100.0	2955	21 AAB18541	Polyprotein encode
39	20	100.0	3011	13 AAR21519	Compiled HCV sequ
40	20	100.0	3011	14 AAR31621	Hepatitis C virus
41	20	100.0	3011	17 AAR90931	Hepatitis C virus
42	20	100.0	3011	18 AAW34480	HCV polypeptin.
43	20	100.0	3011	19 AAW40038	HCV polypeptin.
44	20	100.0	3011	23 AAE22049	Hepatitis C virus
45	20	100.0	3011	23 AAE22052	Hepatitis C virus

ALIGNMENTS

RESULT 1
AAR84505
ID AAR84505 standard; peptide; 20 AA.
AC AAR84505;
XX
DT 06-JAN-1997 (first entry)
XX
Hepatitis C virus peptide NS1-1 (residues 451-470).
DE
Hepatitis C virus; HCV; immunogen; non-structural region; NS1;
KW immunodominant; T cell epitope; vaccine.
XX
Hepatitis C virus.
OS
XX
W09512677-A2.
PN
XX
PD 11-MAY-1995.
XX
PF 28-OCT-1994; 94WO-EP03555.
XX
PR 04-NOV-1993; 93EP-0402718.
XX
(INNO-) INNOGENETICS NV.
PA
Deleys R, Leroux-Roels G, Maertens G;
XX
WPI; 1995-193822/25.
DR
Hepatitis C Virus immunogenic polypeptide contg. a T-cell
PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in
PT production of vaccines, therapeutic agents, etc.

XX Example 4; Page 51; 105pp; English.

PS A series of overlapping peptides (including the present sequence) was synthesised based on sequences in the core, E1 and E2/NS1 regions of hepatitis C virus. The peptides were used as antigens in lymphoproliferative assays to identify the main T-cell epitopes.

CC Sequence 20 AA;

Query Match 100.0%; Score 20; DB 16; Length 20;
 Best Local Similarity 100.0%; Pred. No. 3.2e-14;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCPERLASCRPLTDFDQGWG 20
 |||||
 DB 1 GCPERLASCRPLTDFDQGWG 20

RESULT 2

AAR90999 ID AAR90999 standard; peptide; 20 AA.

XX AC AAR90999;

XX DT 25-SEP-1996 (first entry)

XX DE HCV E2 peptide E2-1B for competition studies.

XX KW HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human; serotype; reversed phase hybridisation assay; genotype; antigen; sera.

XX OS Synthetic.

XX PN WO9604385-A2.

XX PD 15-FEB-1996.

XX PF 31-JUL-1995; 95WO-EP03031.

XX PR 29-JUL-1994; 94EP-0870132.

XX PA (INNO-) INNOGENETICS NV.

XX PI Bosman F, Buyse M, De Martynoff G, Maertens G;

XX WPI; 1996-129401/13.

PT Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope proteins - in presence of disulphide bond cleavage agent, to produce proteins suitable for direct use in vaccines or diagnostic assays of HCV

Example 7; Page 67; 146pp; English.

CC AAR90965-R90980 and AAR90995-R91015 represent synthetic hepatitis C virus (HCV) E1 and E2 peptides used in competition studies. This sequence represents a synthetic E2 peptide, and corresponds to residues 451-470 of the E2 protein sequence. These sequences are useful for in vitro monitoring of HCV disease, or prognosis of the response to interferon treatment of patients suffering from HCV infection. These sequences compete with the proteins produced by AAT12704-T12709 and AAT12961-T12974, which are included in vectors for the production of recombinant E1, E2, and E1/E2. The recombinant proteins can then be isolated and purified by carrying out a disulphide bond cleavage, or a reduction step with a disulphide bond cleavage agent, after lysis of recombinant host cells. The constructs containing the purified HCV envelope proteins can be used for vaccinating humans against HCV, for in vitro detection of HCV antibodies in a sample, and in a serotyping assay for detecting one or more serological types of HCV present in a biological sample. The constructs can also be immobilised on a solid substrate and incorporated into a reversed phase hybridisation assay for determining the presence or the genotype of HCV. The new purification

CC method preserves the conformation of the recombinantly expressed E1, E2 and E1/E2, and eliminates contaminating proteins. Antigens isolated using this method are more reactive with human sera than those isolated by known techniques.

CC Sequence 20 AA;

Query Match 100.0%; Score 20; DB 17; Length 20;
 Best Local Similarity 100.0%; Pred. No. 3.2e-14;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCPERLASCRPLTDFDQGWG 20
 |||||
 DB 1 GCPERLASCRPLTDFDQGWG 20

RESULT 3

AAO18705 ID AAO18705 standard; Peptide; 20 AA.

XX AC AAO18705;

XX DT 24-OCT-2002 (first entry)

XX DE Hepatitis C virus E2 protein derived peptide E2-1B.

XX KW Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide; immunostimulant; vaccine.

XX OS Hepatitis C virus.

XX PN WO200255548-A2.

XX PD 18-JUL-2002.

XX PF 11-JAN-2002; 2002WO-EP00219.

XX PR 11-JAN-2001; 2001US-260699P.

XX PR 30-AUG-2001; 2001US-315768P.

XX PA (INNO-) INNOGENETICS NV.

XX PI Maertens G, Bosman F, Buyse M;

XX WPI; 2002-599657/64.

PT New therapeutic vaccine compositions comprising at least one purified recombinant hepatitis C virus (HCV) single or specific oligomeric recombinant envelope protein E1 or E2, useful for immunizing humans from HCV infection

Example 7; Page 226-227; 243pp; English.

CC The present invention relates to new therapeutic vaccine compositions for inducing hepatitis C virus (HCV) -specific antibodies, comprising a composition containing at least one purified recombinant HCV single or specific oligomeric recombinant envelope proteins selected from an E1 and an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are useful for inducing HCV-specific antibodies or for immunising humans against HCV. The recombinant HCV E1 and/or E2 proteins are useful as vaccines or therapeutics, in HCV screening and confirmatory antibody tests, for raising antibodies in the preparation of medicament, and for in vitro monitoring of HCV disease or prognosis of the response to treatment of patients suffering from HCV infection. The present sequence is a peptide derived from the proteins of the invention.

CC Sequence 20 AA;

Query Match 100.0%; Score 20; DB 23; Length 20;
 Best Local Similarity 100.0%; Pred. No. 3.2e-14;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCPERLASCRPLTDFDQGWG 20

Db 1 GCPERLASCRPLTDFDQGWG 20
|||||

RESULT 4

AAR13359 AAR13359 standard; Protein; 37 AA.

XX AC AAR13359;
XX DT 25-MAR-2003 (updated)
XX DT 23-OCT-1991 (first entry)
XX DE P447 HCV antigen (447-483).
XX KW C100-3; hepatitis C virus; immunoassay; epitope.
XX OS Synthetic.
XX FN AU9068390-A.
XX PD 27-JUN-1991.
XX PF 21-DEC-1990; 90AU-0068390.
XX PR 07-NOV-1990; 90US-0610180.
XX PR 22-DEC-1989; 89US-0456162.
XX PA (ABBO) ABBOTT LAB.
XX PA (LESN/) LESNIEWSKI R R.
XX DR WPI; 1991-238393/33.
XX KW Immunological assays for hepatitis C virus antibody - by using polypeptide(s) of hepatitis C virus antigens

XX Claim 1; Page 45; 62pp; English.

XX The polypeptide may be prepared by solid phase synthesis fragment coupling (pref.) or using recombinant technology.
XX The assay has increased sensitivity and is more specific than assays using the polypeptide C100-3 (EP-318216).
XX See also AAR13146-48 and AAR13343-65.
XX (Updated on 25-MAR-2003 to correct PA field.)
XX SQ Sequence 37 AA;
Query Match 100.0%; Score 20; DB 12; Length 37;
Best Local Similarity 100.0%; Pred. No. 5.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

SQ Sequence 37 AA;

Query Match 100.0%; Score 20; DB 12; Length 37;
Best Local Similarity 100.0%; Pred. No. 5.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
|||||

Db 5 GCPERLASCRPLTDFDQGWG 24

RESULT 5

AAR33878 AAR33878 standard; peptide; 37 AA.

XX AC AAR33878;
XX DT 25-MAR-2003 (updated)
XX DT 19-JUL-1993 (first entry)
XX DE Polypeptide p447 comprising HCV viral antigen.
XX KW Hepatitis C virus; NANBH; assay; antibody; p380-JH1; p380-J; p380LG; p408.
XX OS Synthetic.
XX PN W09306247-A1.

XX 01-APR-1993.
XX PF 16-SEP-1992; 92WO-US07813.
XX PR 16-SEP-1991; 91US-0760292.
XX PA (ABBO) ABBOTT LAB.
XX PI Lesniewski RR, Leung TK;
XX WPI; 1993-117563/14.
XX Assay for detecting presence of antibody to hepatitis C viral antigen - by contacting sample with polypeptide contg. at least one epitope of virus antigen
XX Disclosure; Page 14; 63pp; English.
XX The synthetic peptide p447 represents amino acid residues 447-483 of the hepatitis C viral antigen. The peptide may be used in an assay to detect antibodies to HCV and thus to diagnose chronic HCV infection.
XX See also AAR33861-87.
XX (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 37 AA;

Query Match 100.0%; Score 20; DB 14; Length 37;
Best Local Similarity 100.0%; Pred. No. 5.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
|||||

Db 5 GCPERLASCRPLTDFDQGWG 24

RESULT 6

AAB18528 AAB18528 standard; Protein; 76 AA.

XX AC AAB18528;
XX DT 15-JAN-2001 (first entry)
XX DE Protein encoded by a novel hepatitis C virus cDNA clone 26j.
XX KW Hepatitis C virus; HCV; antisense polynucleotide; polypeptide; viral infectivity; viral replication.
XX OS Hepatitis C virus.
XX PN EP1034785-A2.
XX PD 13-SEP-2000.
XX PF 16-MAR-1990; 2000EP-0109602.
XX PR 17-MAR-1989; 89US-0325338.
XX PR 20-APR-1989; 89US-0341334.
XX PR 18-MAY-1989; 89US-0355002.
XX PR 16-MAR-1990; 90EP-0302866.
XX (CHIR) CHIRON CORP.
XX PI Houghton M, Choo Q, Kuo G;
XX WPI; 2000-566891/53.
XX N-PSDB; AAA75284.
XX Novel composition comprising a hepatitis C virus antisense polynucleotide which is complementary to or corresponds to a sense strand of the virus genome, and selectively hybridises to it -

Example; Fig 4; 75pp; English.

The specification describes a pharmaceutical composition which comprises a hepatitis C virus (HCV) antisense polynucleotide. The HCV is characterized by a positive stranded RNA genome which has 40% homology at the polypeptide level to a HCV polypotein. The antisense polynucleotide binds to cellular polynucleotides which enhance and/or are required for viral infectivity, replicative ability or chronicity. The antisense polynucleotides may also be designed to bind with high specificity, to be of increased stability, to be stable and to have low toxicity. The composition also comprises an agent which causes viral RNA to be inactive. The composition is used for preventing HCV replication in a system. The present sequence is encoded by a novel HCV cDNA sequence, which is used in the course of the invention.

SQ	Sequence	76 AA;
	Query Match	100.0%; Score 20; DB 21; Length 76;
	Best Local Similarity	100.0%; Pred. No. 9.6e-14;
	Matches 20; Conservative	0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
|||
Db 11 GCPERLASCRPLTDFDQGWG 30

RESULT 7
AAB68043
ID AAB68043 standard: protein: 254 AA.

XX	AAB68043;
AC	
XX	
DT	29-JUN-2001 (first entry)

XX Amino acid sequence of water soluble variant of envelope E2 protein
DE
XX
KW Eo protein; HCV; envelope protein; E2 protein; HCV infection;
KW HCV attachment.
KW

AA	
OS	Synthetic.
OS	Hepatitis C virus.
XX	
XX	WO200122984-A1.
PN	
XX	
PD	05-APR-2001.

XX	
PF	26-SEP-2000; 2000WO-US26395.
XX	
PR	29-SEP-1999; 99US-0407430.

XX (UYCO) UNIV COLUMBIA NEW YORK.
PA
XX
PI Worman HJ Mamiya N.

XX
DR WPI; 2001-273486/28.
XX
DT Treating or preventing hepatitis

PT involves administering hepatitis C virus envelope protein E2 binding agents -

XX

PS Claim 5; Fig 8; 46pp; English.

XX

CC The present sequence represents a water soluble variant of a Hepatitis C virus (HCV) envelope E2 protein. Agents that bind to the HCV E2 protein (such as the human Eo protein), and so inhibit the attachment of HCV onto cells (especially liver cells), are used to treat HCV infections in mammals, in particular humans. The specification also describes a method for identifying a compound which can be used for treating or preventing HCV in a subject and which can inhibit the attachment of HCV onto cells by inhibiting the binding of HCV envelope E2 protein to a cellular protein associated with HCV attachment and entry into cells. The method comprises incubating the compound, HCV envelope E2 protein or its variant

CC and a cellular^o protein capable of specifically binding to the HCV E2
CC protein under suitable reaction conditions; determining the interactions
CC between HCV envelope E2 protein and cellular protein in the presence and
CC absence of the compound; and comparing the interaction to identify a
CC compound which can inhibit the attachment of HCV onto cells.

AA	Sequence	254 AA;
SQ		

```

Query Match      100.0%; Score 20; DB 22; Length 254;
Best Local Similarity 100.0%; Pred. No. 2.6e-13;
Matches 20: Conservative 0; Mismatches 0; Indels

```

Qy	1	GC	P	E	R	L	A	S	C	R	P	L	T	D	F	D	Q	G	W	G	20
Dh	46	GC	P	E	R	L	A	S	C	R	P	L	T	D	F	D	Q	G	W	G	65

RESULT 8
AAR33997
ID AAR33997 standard: protein: 278 AA.

XX
AC AAR33997:

XX	
DT	25-MAR-2003 (updated)
DT	26-JUL-1993 (first entry)

Th E2/NS1 protein.

XX Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
 KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
 KW domain; immunological; cross-reactive; envelope protein; vaccine;
 KW gp53 (BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.
 XX
 OS Synthetic.

AA PN WO9306126-A1.

PD 01-APR-1993.

PF 11-SEP-1992; 92WO-US07683.

PR 13-SEP-1991; 91US-0759575.

PA (CHIR) CHIRON CORP.

PI Houghton M, Weiner AJ;

DR WPI; 1993-117468/14.

Immuo-reactive hepatitis C virus polypeptide compsns. - contg.
at least 2 sequences from the first variable domain of distinct
HCV isolates

XX PS Disclosure: Fig 3: 106pp; English.

The sequences given in AAR33992-002 represent a portion of the E2/NS1 protein encoded by group I and group II HCV isolates, from amino acid 370-850. E2/NS1 are viral envelope proteins and are of immunogenic interest. E2/NS1 contains an N-terminal hypervariable domain of about 30 amino acids which shows large variation between nearly all isolates. This is an important immunoreactive domain. This putative envelope glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera virus) envelope polypeptide of the pestiviruses and the NS1 of the flaviviruses, both of which confer protective immunity in hosts vaccinated with these polypeptides. It has been discovered that a number of important HCV epitopes vary among viral isolates and that these epitopes can be mapped to specific domains. This meant that immunologically cross-reactive polypeptides which focus on variable rather than constant domains can be produced. See also AAQ33134-48 and AAR33982-91.

(Updated on 25-MAR-2003 to correct PN field.)

CC and ARK33982-51.
CC (Updated on 25-MAR-2003 to correct PN field.)

AA	Sequence	278 AA;
SQ		

Query Match 100.0%; Score 20; DB 14; Length 278;
 Best Local Similarity 100.0%; Pred. No. 2.8e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCPERLASCRPLTDFDQGWG 20
 DB 82 GCPERLASCRPLTDFDQGWG 101

RESULT 9
 AAB68042
 ID AAB68042 standard; protein; 363 AA.
 AC AAB68042;
 XX 29-JUN-2001 (first entry)
 DT
 DE Amino acid sequence of a Hepatitis C virus envelope E2 protein.
 KW Eo protein; HCV; envelope protein; E2 protein; HCV infection;
 KW HCV attachment.
 XX Hepatitis C virus.
 OS
 XX WO200122984-A1.
 PN
 XX 05-APR-2001.
 PD
 XX 26-SEP-2000; 2000WO-US26395.
 PF
 XX 29-SEP-1999; 99US-0407430.
 PR
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 PA
 XX Worman HJ, Mamiya N;
 PI
 XX WPI; 2001-273486/28.
 DR
 XX Treating or preventing hepatitis C virus infection in a subject.
 PT involves administering hepatitis C virus envelope protein E2 binding
 PT agents -
 XX
 PS Claim 3; Fig 7; 46pp; English.
 CC The present sequence represents a Hepatitis C virus (HCV) envelope
 CC E2 protein. Agents that bind to the HCV E2 protein (such as the human
 CC Eo protein), and so inhibit the attachment of HCV onto cells
 CC (especially liver cells), are used to treat HCV infections in mammals,
 CC in particular humans. The specification also describes a method for
 CC identifying a compound which can be used for treating or preventing
 CC HCV in a subject and which can inhibit the attachment of HCV onto cells
 CC by inhibiting the binding of HCV envelope E2 protein to a cellular
 CC protein associated with HCV attachment and entry into cells. The method
 CC comprises incubating the compound, HCV envelope E2 protein or its variant
 CC and a cellular protein capable of specifically binding to the HCV E2
 CC protein under suitable reaction conditions; determining the interactions
 CC between HCV envelope E2 protein and cellular protein in the presence and
 CC absence of the compound; and comparing the interaction to identify a
 CC compound which can inhibit the attachment of HCV onto cells.
 XX
 SQ Sequence 363 AA;

Query Match 100.0%; Score 20; DB 22; Length 363;
 Best Local Similarity 100.0%; Pred. No. 3.5e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCPERLASCRPLTDFDQGWG 20
 DB 68 GCPERLASCRPLTDFDQGWG 87

RESULT 10

AAR34438
 ID AAR34438 standard; Protein; 402 AA.
 XX
 AC AAR34438;
 XX 25-MAR-2003 (updated)
 DT 09-AUG-1993 (first entry)
 XX
 DE Sequence of glycoprotein E2/NS1 in clone HCV1.
 XX
 KW Hepatitis C virus; envelope protein; glycoprotein; E2/NS1;
 KW diagnostic reagent.
 XX
 OS Hepatitis C virus.
 XX
 PN EP537626-A1.
 XX
 PD 21-APR-1993.
 XX
 PF 08-OCT-1992; 92EP-0117191.
 XX
 PR 08-OCT-1991; 91JP-0260824.
 XX
 PA (NAHE-) NAT INST OF HEALTH.
 XX
 PI Harada S, Honda Y, Miyamura T, Saito I;
 XX
 DR WPI; 1993-127516/16.
 DR N-PSDB; AAQ40330.
 XX
 PT Diagnostic reagent for hepatitis C virus - comprises second
 PT envelope protein or first non-structural protein encoded by HCV
 PT gene and has sugar chain
 XX
 PS Claim 2; Pages 30-32; 58pp; English.
 XX
 CC Glycoprotein E2/NS1 is derived from the second envelope protein or
 CC first non-structural protein encoded by the genome of HCV. The
 CC nucleic acid is extracted from the serum of the patient of hepatitis
 CC C. The serum is pref. mixed with transfer RNA (tRNA) as a carrier
 CC of virus RNA. As a technique of cloning cDNA from the nucleic acid,
 CC it is preferred to use polymerase chain reaction method. In the
 CC reaction, any commercially available random primers or synthesized
 CC DNA having a base sequence similar to that of primer AS1 may be used
 CC as a primer. Representative examples of sense primers include S1.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 402 AA;

Query Match 100.0%; Score 20; DB 14; Length 402;
 Best Local Similarity 100.0%; Pred. No. 3.8e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCPERLASCRPLTDFDQGWG 20
 DB 112 GCPERLASCRPLTDFDQGWG 131

RESULT 11
 AAP90183
 ID AAP90183 standard; protein; 454 AA.
 XX
 AC AAP90183;
 XX 25-MAR-2003 (updated)
 DT 01-NOV-1989 (first entry)
 XX
 DE Sequence of hepatitis C virus cDNA insert in clone k9-1.
 XX
 KW Hepatitis C virus; clone k9-1; probe; vaccine.
 XX
 OS Pan troglodytes.
 XX

XX The specification describes a pharmaceutical composition which
 CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
 CC HCV is characterized by a positive stranded RNA genome which has
 CC 40% homology at the polypeptide level to a HCV polypeptide. The
 CC antisense polynucleotide binds to cellular polynucleotides which
 CC enhance and/or are required for viral infectivity, replicative
 CC ability or chronicity. The antisense polynucleotides may also be
 CC designed to bind with high specificity, to be of increased stability,
 CC to be stable and to have low toxicity. The composition also comprises
 CC an agent which causes viral RNA to be inactive. The composition
 CC is used for preventing HCV replication in a system. The present
 CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
 CC course of the invention.

XX SQ Sequence 454 AA;
 CC Query Match 100.0%; Score 20; DB 21; Length 454;
 CC Best Local Similarity 100.0%; Pred. No. 4.2e-13;
 CC Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
 |||||
 Db 1 GCPERLASCRPLTDFDQGWG 20

RESULT 14
 AAR33588
 ID AAR33588 standard; protein; 463 AA.
 XX AAR33588;
 AC AAR33588;
 XX AAR33588;
 XX AAR33588;
 DT 25-MAR-2003 (updated)
 DT 05-JUL-1993 (first entry)
 XX HCV CKS-NS1S1 fusion antigen.
 DE HCV CKS-NS1S1 fusion antigen.
 XX Hepatitis C Virus; non-A, non-B hepatitis virus; NANBH;
 KW non-structural protein; CMP-KDO synthetase; CKS fusion protein;
 KW CTP: CMP-3-deoxy-manno-octulosonate cytidyl transferase;
 KW immunoassay; pHCV-77.
 XX Hepatitis C Virus.
 OS Hepatitis C Virus.
 XX WO9304088-A1.
 XX 04-MAR-1993.
 XX 21-AUG-1992; 92WO-US07188.
 XX 21-AUG-1991; 91US-0748561.
 XX (ABBO) ABBOTT LAB.
 XX Dailey SH, Desai SM, Devare SG;
 XX WPI; 1993-093941/11.
 XX Hepatitis C assay using recombinant NS1 region antigens - for
 PT detecting antibodies and antigen in body fluids from individuals
 PT exposed to hepatitis C virus
 XX Claim 1; Page 38-40; 175pp; English.
 XX Eight oligonucleotides representing amino acids 365-579 of the HCV
 CC genome were ligated together and cloned as a 645bp EcoRI/BamHI
 CC fragment into the CKS fusion vector pJO200. The amino acid sequence
 CC of this antigen is designated pHCV-77 (i.e. AAR33588). The resultant
 CC fusion protein HCV CKS-NS1S1 consists of 239 amino acids of CKS,
 CC seven amino acids contributed by linker DNA sequences and 215 amino
 CC acids from the NS1 region of the HCV genome. The fusion protein is
 CC used to detect antibodies and antigen in body fluids from
 CC individuals exposed to HCV.

CC (Updated on 25-MAR-2003 to correct PN field.)
 XX SQ Sequence 463 AA;
 CC Query Match 100.0%; Score 20; DB 14; Length 463;
 CC Best Local Similarity 100.0%; Pred. No. 4.2e-13;
 CC Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
 |||||
 Db 335 GCPERLASCRPLTDFDQGWG 354

RESULT 15
 AAR33187
 ID AAR33187 standard; peptide; 463 AA.
 XX AAR33187;
 AC AAR33187;
 XX AAR33187;
 DT 25-MAR-2003 (updated)
 DT 03-JUL-1993 (first entry)
 XX Sequence of subfragment pHCV65 (AAs 565-731) of the hepatitis
 DE C virus (HCV) genome.
 DE C virus (HCV) genome.
 XX Immunogenic peptide; hepatitis C virus; immunogenic domain;
 KW monoclonal antibody; diagnosis; detection; therapy.
 XX Synthetic.
 XX WO9304205-A1.
 XX 04-MAR-1993.
 XX 21-AUG-1992; 92WO-US07189.
 XX 21-AUG-1991; 91US-0748292.
 XX (ABBO) ABBOTT LAB.
 XX Dailey SH, Desai SM, Devare SG, Johnson JE, Mehta SU;
 XX WPI; 1993-094030/11.
 XX Monoclonal antibodies specific for hepatitis C virus E2-NS1
 PT antigen - useful for diagnosis and evaluation of HCV infections
 PT and in differentiation studies
 XX Example; Pages 36-37; 48pp; English.
 XX Immunogenic domains of E2/NS1 region of HCV genome encompassing AAs
 CC 600-720 were mapped with PEPSKAN analysis. Based on the EIA reactivity
 CC of a panel of HCV positive sera, peptide AAR33184 was chosen as the
 CC immunogen for the generation of monoclonal antibodies to HCV NS1.
 CC Several individual oligos representing AAs 365-731 of HCV genome
 CC were ligated and cloned as three separate EcoRI/BamHI subfragments
 CC into the CKS fusion vector pJO200. The sequences of these three
 CC subfragments - pHCV80, pHCV77 and pHCV65 are in AAR33185-7.
 CC Analysis showed that Mabs H13C113 and H23C163 showed reactivity
 CC pHCV80 and pHCV65.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX SQ Sequence 463 AA;
 CC Query Match 100.0%; Score 20; DB 14; Length 463;
 CC Best Local Similarity 100.0%; Pred. No. 4.2e-13;
 CC Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
 |||||
 Db 335 GCPERLASCRPLTDFDQGWG 354

Search completed: November 21, 2003, 20:58:03
Job time : 32.15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.6 Seconds
(without alignments)
79.832 Million cell updates/sec

Title: US-09-973-025-76

Perfect score: 20

Sequence: 1 GCPERLASCRPLTDFDQGWG 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pap.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-08-612-973-76
2	20	100.0	20	3	US-08-927-597-76
3	20	100.0	20	4	US-08-635-886C-32
4	20	100.0	37	5	PCT-US92-07813-18
5	20	100.0	76	3	US-08-444-818-79
6	20	100.0	278	1	US-08-440-103-15
7	20	100.0	278	1	US-08-440-542-15
8	20	100.0	278	1	US-08-231-368-15
9	20	100.0	278	4	US-08-440-210-15
10	20	100.0	278	4	US-09-046-604-15
11	20	100.0	402	1	US-08-460-806-13
12	20	100.0	402	1	US-08-325-630-13
13	20	100.0	403	2	US-08-483-695-39
14	20	100.0	403	2	US-07-965-285-39
15	20	100.0	403	2	US-08-487-231-39
16	20	100.0	403	3	US-09-201-912-39
17	20	100.0	454	3	US-08-444-818-73
18	20	100.0	463	1	US-07-748-292-9
19	20	100.0	463	3	US-08-867-611-31
20	20	100.0	463	5	PCT-US92-06965A-1
21	20	100.0	480	1	US-08-440-103-14
22	20	100.0	480	1	US-08-440-542-14
23	20	100.0	480	1	US-08-231-368-14
24	20	100.0	480	1	US-08-440-210-14
25	20	100.0	480	4	US-09-046-604-14
26	20	100.0	621	1	US-07-748-292-7
27	20	100.0	622	3	US-08-867-611-34

28	20	100.0	622	5	PCT-US92-06965A-4	Sequence 4, Appli
29	20	100.0	663	3	US-08-824-057-3	Sequence 3, Appli
30	20	100.0	663	4	US-09-415-582-3	Sequence 3, Appli
31	20	100.0	663	4	US-09-693-596-4	Sequence 4, Appli
32	20	100.0	663	3	US-08-867-611-35	Sequence 35, Appli
33	20	100.0	738	5	PCT-US92-06965A-5	Sequence 5, Appli
34	20	100.0	2436	3	US-08-444-818-75	Sequence 75, Appli
35	20	100.0	2772	3	US-08-444-818-89	Sequence 89, Appli
36	20	100.0	2894	2	US-08-466-975A-23	Sequence 23, Appli
37	20	100.0	2894	2	US-08-391-671A-23	Sequence 23, Appli
38	20	100.0	2894	3	US-08-467-902A-23	Sequence 23, Appli
39	20	100.0	2894	3	US-09-275-285-23	Sequence 23, Appli
40	20	100.0	2894	4	US-09-941-611-23	Sequence 23, Appli
41	20	100.0	2955	2	US-08-443-260-3	Sequence 3, Appli
42	20	100.0	2955	3	US-08-442-805A-3	Sequence 3, Appli
43	20	100.0	2955	3	US-08-443-900A-3	Sequence 3, Appli
44	20	100.0	2955	3	US-08-444-818-124	Sequence 124, App
45	20	100.0	2955	3	US-08-249-843-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-612-973-76
; Sequence 76, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612.973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-612-973-76

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCPERLASCRPLTDFDQGWG 20
|||||

Db 1 GCPERLASCRPLTDFDQGWG 20

RESULT 2

US-08-927-597-76
; Sequence 76, Application US/08927597
; Patent No. 6245503
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927.597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-927-597-76

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20

Db 1 GCPERLASCRPLTDFDQGWG 20

RESULT 3

US-08-635-886C-32
; Sequence 32, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEUX, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555

; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-32

Query Match 100.0%; Score 20; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20

Db 1 GCPERLASCRPLTDFDQGWG 20

RESULT 4

PCT-US92-07813-18
; Sequence 18, Application PC/TUS9207813
; GENERAL INFORMATION:
; APPLICANT: LESNIEWSKI, RICHARD R.
; APPLICANT: LEUNG, TAT K.
; TITLE OF INVENTION: HEPATITIS C ASSAY
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: U.S.A.
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/07813
; FILING DATE: 19920916
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKIP, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4767.P3.03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 37 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US92-07813-18

Query Match 100.0%; Score 20; DB 5; Length 37;
Best Local Similarity 100.0%; Pred. No. 2.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20

Db 5 GCPERLASCRPLTDFDQGWG 24

RESULT 5

US-08-444-818-79
; Sequence 79, Application US/08444818
; Patent No. 6150087

;; GENERAL INFORMATION:
;; APPLICANT: Chien, David Y.
;; TITLE OF INVENTION: NABV Diagnostics and Vaccines
;; NUMBER OF SEQUENCES: 777
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Chiron Corporation
;; STREET: 4560 Horton Street
;; CITY: Emeryville
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94608-2916
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/444,818
;; FILING DATE:
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/403,590
;; FILING DATE: 14-MAR-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Harbin, Aliea A.
;; REGISTRATION NUMBER: 33,895
;; REFERENCE/DOCKET NUMBER: 0110.002
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (508)359-3876
;; TELEFAX: (508)359-3885
;; INFORMATION FOR SEQ ID NO: 79:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 76 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-444-818-79

Query Match 100.0%; Score 20; DB 3; Length 76;
Best Local Similarity 100.0%; Pred. No. 3.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
Db 11 GCPERLASCRPLTDFDQGWG 30

RESULT 6
US-08-440-103-15
; Sequence 15, Application US/08440103
; Patent No. 5670152
;; GENERAL INFORMATION:
;; APPLICANT: Weiner, Amy J.
;; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
;; NUMBER OF SEQUENCES: 45
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Chiron Corporation
;; STREET: 4560 Horton Street
;; CITY: Emeryville
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94608
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/440,103
;; FILING DATE: 12-MAY-1995
;; CLASSIFICATION: 424

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/231,368
;; FILING DATE:
;; APPLICATION NUMBER: US 07/759,575
;; FILING DATE: 13-SEP-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: McClung, Barbara G.
;; REGISTRATION NUMBER: 33,113
;; REFERENCE/DOCKET NUMBER: 0205.001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (510) 601-2708
;; TELEFAX: (510) 655-3542
;; INFORMATION FOR SEQ ID NO: 15:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 278 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-440-103-15

Query Match 100.0%; Score 20; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
Db 82 GCPERLASCRPLTDFDQGWG 101

RESULT 7
US-08-440-542-15
; Sequence 15, Application US/08440542
; Patent No. 5670153
;; GENERAL INFORMATION:
;; APPLICANT: Weiner, Amy J.
;; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
;; NUMBER OF SEQUENCES: 45
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Chiron Corporation
;; STREET: 4560 Horton Street
;; CITY: Emeryville
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94608
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/440,542
;; FILING DATE: 12-MAY-1995
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/231,368
;; FILING DATE:
;; APPLICATION NUMBER: US 07/759,575
;; FILING DATE: 13-SEP-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: McClung, Barbara G.
;; REGISTRATION NUMBER: 33,113
;; REFERENCE/DOCKET NUMBER: 0205.001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (510) 601-2708
;; TELEFAX: (510) 655-3542
;; INFORMATION FOR SEQ ID NO: 15:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 278 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear

```
; MOLECULE TYPE: protein
US-08-440-542-15

Query Match      100.0%; Score 20; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GCPERLASCRPLTDFDQGWG 20
      |||||
Db      82 GCPERLASCRPLTDFDQGWG 101

RESULT 8
US-08-231-368-15
; Sequence 15, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/231.368
; FILING DATE: 13-SEP-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-231-368-15

Query Match      100.0%; Score 20; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GCPERLASCRPLTDFDQGWG 20
      |||||
Db      82 GCPERLASCRPLTDFDQGWG 101

RESULT 9
US-08-440-210-15
; Sequence 15, Application US/08440210
; Patent No. 5766845
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
```

```
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/440,210
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-210-15

Query Match      100.0%; Score 20; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GCPERLASCRPLTDFDQGWG 20
      |||||
Db      82 GCPERLASCRPLTDFDQGWG 101

RESULT 10
US-09-046-604-15
; Sequence 15, Application US/09046604
; Patent No. 6303292
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/09/046,604
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-046-604-15

Query Match      100.0%; Score 20; DB 4; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCPERLASCRPLTDFDQGWG 20
DB      82 GCPERLASCRPLTDFDQGWG 101

RESULT 11
US-08-460-806-13
; Sequence 13, Application US/08460806
; Patent No. 5747241
; GENERAL INFORMATION:
; APPLICANT: MIYAMURA, TATSUO
; APPLICANT: SAITO, IZUMU
; APPLICANT: HARADA, SHIZUKO
; APPLICANT: HONDA, YOSHIKAZU
; TITLE OF INVENTION: DIAGNOSTIC REAGENT FOR HEPATITIS C
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCELLELAND, MAIER & NEUSTADT,
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,806
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,630
; FILING DATE: 19-OCT-1994
; APPLICATION NUMBER: US 07/956,993
; FILING DATE: 06-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5747241man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 4667-001-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-046-604-15

Query Match      100.0%; Score 20; DB 4; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCPERLASCRPLTDFDQGWG 20
DB      82 GCPERLASCRPLTDFDQGWG 101
```

```
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-460-806-13

Query Match      100.0%; Score 20; DB 1; Length 402;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCPERLASCRPLTDFDQGWG 20
DB      112 GCPERLASCRPLTDFDQGWG 131

RESULT 12
US-08-325-630-13
; Sequence 13, Application US/08325630
; Patent No. 5750331
; GENERAL INFORMATION:
; APPLICANT: MIYAMURA, TATSUO
; APPLICANT: SAITO, IZUMU
; APPLICANT: HARADA, SHIZUKO
; APPLICANT: HONDA, YOSHIKAZU
; TITLE OF INVENTION: DIAGNOSTIC REAGENT FOR HEPATITIS C
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCELLELAND, MAIER & NEUSTADT,
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,630
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/956,993
; FILING DATE: 06-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5750331man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 4667-001-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-325-630-13

Query Match      100.0%; Score 20; DB 1; Length 402;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCPERLASCRPLTDFDQGWG 20
DB      112 GCPERLASCRPLTDFDQGWG 131

RESULT 13
US-08-483-695-39
; Sequence 39, Application US/08483695
```

Patent No. 5866139
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Kremendorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; TITLE OF INVENTION: Applications
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483.695
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/965.285
; FILING DATE: 18-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 05286-0001-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4400
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-483-695-39
Query Match 100.0%; Score 20; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCPERLASCRPLTDFDQGWG 20
DB 118 GCPERLASCRPLTDFDQGWG 137
RESULT 14
US-07-965-285-39
; Sequence 39, Application US/07965285
; Patent No. 5879904
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Kremendorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; TITLE OF INVENTION: Applications
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC

COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/965.285
; FILING DATE: 18-MAR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 06 882
; FILING DATE: 06-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 05286-0001-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-965-285-39
Query Match 100.0%; Score 20; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCPERLASCRPLTDFDQGWG 20
DB 118 GCPERLASCRPLTDFDQGWG 137
RESULT 15
US-08-487-231-39
; Sequence 39, Application US/08487231
; Patent No. 5919454
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Kremendorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; TITLE OF INVENTION: Applications
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487.231
; FILING DATE: 07-JUNE-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/965,285
; FILING DATE: 18-MAR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 06 882

; FILING DATE: 06-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 05286-0001-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-487-231-39

Query Match 100.0%; Score 20; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCPERLASCRPLTDFDQGWG 20
Db 118 GCPERLASCRPLTDFDQGWG 137

Search completed: November 21, 2003, 21:15:15
Job time : 11.6 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time 19.5 Seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-76

Perfect score: 20

Sequence: 1 GCPERLASCRPLTDFDQGWG 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US05_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	10	US-09-973-025-76
2	20	100.0	20	11	US-09-899-303-76
3	20	100.0	20	11	US-09-995-808-76
4	20	100.0	20	11	US-09-995-860-76
5	20	100.0	20	12	US-09-995-791-76
6	20	100.0	254	10	US-09-407-430-3
7	20	100.0	363	10	US-09-407-430-2
8	20	100.0	637	12	US-10-187-257-4
9	20	100.0	637	12	US-10-265-083-2
10	20	100.0	2894	10	US-09-941-611-23
11	20	100.0	2894	15	US-10-044-995-23
12	20	100.0	3011	9	US-09-916-353-2
13	20	100.0	3011	16	US-10-232-643-6
14	10	50.0	38	16	US-10-318-200-10
15	10	50.0	91	10	US-09-921-397-80

16	10	50.0	250	10	US-09-952-572-8	Sequence 8, Appli
17	10	50.0	290	12	US-10-128-587A-3	Sequence 3, Appli
18	10	50.0	290	15	US-10-128-590-3	Sequence 3, Appli
19	10	50.0	301	12	US-10-128-587A-5	Sequence 5, Appli
20	10	50.0	301	15	US-10-128-590-5	Sequence 5, Appli
21	10	50.0	314	10	US-09-973-025-42	Sequence 42, Appli
22	10	50.0	314	11	US-09-899-303-42	Sequence 42, Appli
23	10	50.0	314	11	US-09-995-808-42	Sequence 42, Appli
24	10	50.0	314	11	US-09-995-860-42	Sequence 42, Appli
25	10	50.0	314	12	US-09-995-791-42	Sequence 42, Appli
26	10	50.0	319	10	US-09-973-025-44	Sequence 44, Appli
27	10	50.0	319	11	US-09-899-303-44	Sequence 44, Appli
28	10	50.0	319	11	US-09-995-808-44	Sequence 44, Appli
29	10	50.0	319	11	US-09-995-860-44	Sequence 44, Appli
30	10	50.0	338	10	US-09-995-791-44	Sequence 38, Appli
31	10	50.0	338	10	US-09-973-025-38	Sequence 38, Appli
32	10	50.0	338	11	US-09-899-303-38	Sequence 38, Appli
33	10	50.0	338	11	US-09-995-808-38	Sequence 38, Appli
34	10	50.0	338	11	US-09-995-860-38	Sequence 38, Appli
35	10	50.0	343	10	US-09-973-025-40	Sequence 40, Appli
36	10	50.0	343	11	US-09-899-303-40	Sequence 40, Appli
37	10	50.0	343	11	US-09-995-808-40	Sequence 40, Appli
38	10	50.0	343	11	US-09-995-860-40	Sequence 40, Appli
39	10	50.0	350	10	US-09-929-955-4	Sequence 4, Appli
40	10	50.0	350	14	US-10-104-966-4	Sequence 4, Appli
41	10	50.0	363	12	US-10-128-587A-97	Sequence 97, Appli
42	10	50.0	363	12	US-10-128-587A-98	Sequence 98, Appli
43	10	50.0	363	12	US-10-128-587A-97	Sequence 97, Appli
44	10	50.0	363	15	US-10-128-590-97	Sequence 97, Appli
45	10	50.0	363	15	US-10-128-590-97	Sequence 97, Appli

ALIGNMENTS

RESULT 1

US-09-973-025-76
; Sequence 76, Application US/09973025
; Publication No. US20020182706A1
; GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

BOSMAN, FONS

DE MARTYNOFF, GUY

BUYSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHVE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLYING APPLICATION NUMBER: US/09/973,025

FILING DATE: 10-Oct-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 76:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 76:
US-09-973-025-76

Query Match 100.0%; Score 20; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
Db 1 GCPERLASCRPLTDFDQGWG 20

RESULT 2

US-09-899-303-76
; Sequence 76, Application US/09899303
; Publication No. US20030036110A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 76:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 76:
US-09-899-303-76

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
Db 1 GCPERLASCRPLTDFDQGWG 20

Db 1 GCPERLASCRPLTDFDQGWG 20

RESULT 3

US-09-995-808-76
; Sequence 76, Application US/09995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 76
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-76

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
Db 1 GCPERLASCRPLTDFDQGWG 20

RESULT 4

US-09-995-860-76
; Sequence 76, Application US/09995860
; Publication No. US20030118603A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 76
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-76

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
Db 1 GCPERLASCRPLTDFDQGWG 20

RESULT 5

US-09-995-791-76
; Sequence 76, Application US/09995791
; Publication No. US20030147918A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 76


```
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-76

Query Match      100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
    |||||
Db 1 GCPERLASCRPLTDFDQGWG 20
    |||||

RESULT 6
US-09-407-430-3
; Sequence 3, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; TITLE OF INVENTION: HEPATITIS C VIRUS INFECTION
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-407-430-3

Query Match      100.0%; Score 20; DB 10; Length 254;
Best Local Similarity 100.0%; Pred. No. 1.9e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
    |||||
Db 46 GCPERLASCRPLTDFDQGWG 65
    |||||

RESULT 7
US-09-407-430-2
; Sequence 2, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; TITLE OF INVENTION: HEPATITIS C VIRUS INFECTION
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-09-407-430-2

Query Match      100.0%; Score 20; DB 10; Length 363;
Best Local Similarity 100.0%; Pred. No. 2.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
    |||||
Db 68 GCPERLASCRPLTDFDQGWG 87
    |||||

RESULT 8
US-10-187-257-4
; Sequence 4, Application US/10187257
; Publication No. US20030138458A1
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: O'HAGAN, Derek
; TITLE OF INVENTION: HCV E1E2 VACCINE COMPOSITIONS
; FILE REFERENCE: 2302-17206
; CURRENT APPLICATION NUMBER: US/10/187,257
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
US-10-187-257-4

Query Match      100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 4.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
    |||||
Db 279 GCPERLASCRPLTDFDQGWG 298
    |||||

RESULT 9
US-10-265-083-2
; Sequence 2, Application US/10265083
; Publication No. US20030170273A1
; GENERAL INFORMATION:
; APPLICANT: VALIANTE, Nicholas
; APPLICANT: O'HAGAN, Derek
; TITLE OF INVENTION: ADJUVANT COMPOSITIONS
; FILE REFERENCE: 2302-18357.30
; CURRENT APPLICATION NUMBER: US/10/265,083
; CURRENT FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7
US-10-265-083-2

Query Match      100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 4.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
    |||||
Db 279 GCPERLASCRPLTDFDQGWG 298
    |||||

RESULT 10
US-09-941-611-23
; Sequence 23, Application US/09941611
; Patent No. US20020106640A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSWUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; ANTIBODIES TO HEPATITIS C VIRUS
```

1 ;
2 ;
3 ;
4 ;
5 ;
6 ;
7 ;
8 ;
9 ;
10 ;
11 ;
12 ;
13 ;
14 ;
15 ;
16 ;
17 ;
18 ;
19 ;
20 ;
21 ;
22 ;
23 ;
24 ;
25 ;
26 ;
27 ;
28 ;
29 ;
30 ;
31 ;
32 ;
33 ;
34 ;
35 ;
36 ;
37 ;
38 ;
39 ;
40 ;
41 ;
42 ;
43 ;
44 ;
45 ;
46 ;
47 ;
48 ;
49 ;
50 ;
51 ;
52 ;
53 ;
54 ;
55 ;
56 ;
57 ;
58 ;
59 ;
60 ;
61 ;
62 ;
63 ;
64 ;
65 ;
66 ;
67 ;
68 ;
69 ;
70 ;
71 ;
72 ;
73 ;
74 ;
75 ;
76 ;
77 ;
78 ;
79 ;
80 ;
81 ;
82 ;
83 ;
84 ;
85 ;
86 ;
87 ;
88 ;
89 ;
90 ;
91 ;
92 ;
93 ;
94 ;
95 ;
96 ;
97 ;
98 ;
99 ;
100 ;
101 ;
102 ;
103 ;
104 ;
105 ;
106 ;
107 ;
108 ;
109 ;
110 ;
111 ;
112 ;
113 ;
114 ;
115 ;
116 ;
117 ;
118 ;
119 ;
120 ;
121 ;
122 ;
123 ;
124 ;
125 ;
126 ;
127 ;
128 ;
129 ;
130 ;
131 ;
132 ;
133 ;
134 ;
135 ;
136 ;
137 ;
138 ;
139 ;
140 ;
141 ;
142 ;
143 ;
144 ;
145 ;
146 ;
147 ;
148 ;
149 ;
150 ;
151 ;
152 ;
153 ;
154 ;
155 ;
156 ;
157 ;
158 ;
159 ;
160 ;
161 ;
162 ;
163 ;
164 ;
165 ;
166 ;
167 ;
168 ;
169 ;
170 ;
171 ;
172 ;
173 ;
174 ;
175 ;
176 ;
177 ;
178 ;
179 ;
180 ;
181 ;
182 ;
183 ;
184 ;
185 ;
186 ;
187 ;
188 ;
189 ;
190 ;
191 ;
192 ;
193 ;
194 ;
195 ;
196 ;
197 ;
198 ;
199 ;
200 ;
201 ;
202 ;
203 ;
204 ;
205 ;
206 ;
207 ;
208 ;
209 ;
210 ;
211 ;
212 ;
213 ;
214 ;
215 ;
216 ;
217 ;
218 ;
219 ;
220 ;
221 ;
222 ;
223 ;
224 ;
225 ;
226 ;
227 ;
228 ;
229 ;
230 ;
231 ;
232 ;
233 ;
234 ;
235 ;
236 ;
237 ;
238 ;
239 ;
240 ;
241 ;
242 ;
243 ;
244 ;
245 ;
246 ;
247 ;
248 ;
249 ;
250 ;
251 ;
252 ;
253 ;
254 ;
255 ;
256 ;
257 ;
258 ;
259 ;
260 ;
261 ;
262 ;
263 ;
264 ;
265 ;
266 ;
267 ;
268 ;
269 ;
270 ;
271 ;
272 ;
273 ;
274 ;
275 ;
276 ;
277 ;
278 ;
279 ;
280 ;
281 ;
282 ;
283 ;
284 ;
285 ;
286 ;
287 ;
288 ;
289 ;
290 ;
291 ;
292 ;
293 ;
294 ;
295 ;
296 ;
297 ;
298 ;
299 ;
300 ;
301 ;
302 ;
303 ;
304 ;
305 ;
306 ;
307 ;
308 ;
309 ;
310 ;
311 ;
312 ;
313 ;
314 ;
315 ;
316 ;
317 ;
318 ;
319 ;
320 ;
321 ;
322 ;
323 ;
324 ;
325 ;
326 ;
327 ;
328 ;
329 ;
330 ;
331 ;
332 ;
333 ;
334 ;
335 ;
336 ;
337 ;
338 ;
339 ;
340 ;
341 ;
342 ;
343 ;
344 ;
345 ;
346 ;
347 ;
348 ;
349 ;
350 ;
351 ;
352 ;
353 ;
354 ;
355 ;
356 ;
357 ;
358 ;
359 ;
360 ;
361 ;
362 ;
363 ;
364 ;
365 ;
366 ;
367 ;
368 ;
369 ;
370 ;
371 ;
372 ;
373 ;
374 ;
375 ;
376 ;
377 ;
378 ;
379 ;
380 ;
381 ;
382 ;
383 ;
384 ;
385 ;
386 ;
387 ;
388 ;
389 ;
390 ;
391 ;
392 ;
393 ;
394 ;
395 ;
396 ;
397 ;
398 ;
399 ;
400 ;
401 ;
402 ;
403 ;
404 ;
405 ;
406 ;
407 ;
408 ;
409 ;
410 ;
411 ;
412 ;
413 ;
414 ;
415 ;
416 ;
417 ;
418 ;
419 ;
420 ;
421 ;
422 ;
423 ;
424 ;
425 ;
426 ;
427 ;
428 ;
429 ;
430 ;
431 ;
432 ;
433 ;
434 ;
435 ;
436 ;
437 ;
438 ;
439 ;
440 ;
441 ;
442 ;
443 ;
444 ;
445 ;
446 ;
447 ;
448 ;
449 ;
450 ;
451 ;
452 ;
453 ;
454 ;
455 ;
456 ;
457 ;
458 ;
459 ;
460 ;
461 ;
462 ;
463 ;
464 ;
465 ;
466 ;
467 ;
468 ;
469 ;
470 ;
471 ;
472 ;
473 ;
474 ;
475 ;
476 ;
477 ;
478 ;
479 ;
480 ;
481 ;
482 ;
483 ;
484 ;
485 ;
486 ;
487 ;
488 ;
489 ;
490 ;
491 ;
492 ;
493 ;
494 ;
495 ;
496 ;
497 ;
498 ;
499 ;
500 ;
501 ;
502 ;
503 ;
504 ;
505 ;
506 ;
507 ;
508 ;
509 ;
510 ;
511 ;
512 ;
513 ;
514 ;
515 ;
516 ;
517 ;
518 ;
519 ;
520 ;
521 ;
522 ;
523 ;
524 ;
525 ;
526 ;
527 ;
528 ;
529 ;
530 ;
531 ;
532 ;
533 ;
534 ;
535 ;
536 ;
537 ;
538 ;
539 ;
540 ;
541 ;
542 ;
543 ;
544 ;
545 ;
546 ;
547 ;
548 ;
549 ;
550 ;
551 ;
552 ;
553 ;
554 ;
555 ;
556 ;
557 ;
558 ;
559 ;
560 ;
561 ;
562 ;
563 ;
564 ;
565 ;
566 ;
567 ;
568 ;
569 ;
570 ;
571 ;
572 ;
573 ;
574 ;
575 ;
576 ;
577 ;
578 ;
579 ;
580 ;
581 ;
582 ;
583 ;
584 ;
585 ;
586 ;
587 ;
588 ;
589 ;
590 ;
591 ;
592 ;
593 ;
594 ;
595 ;
596 ;
597 ;
598 ;
599 ;
600 ;
601 ;
602 ;
603 ;
604 ;
605 ;
606 ;
607 ;
608 ;
609 ;
610 ;
611 ;
612 ;
613 ;
614 ;
615 ;
616 ;
617 ;
618 ;
619 ;
620 ;
621 ;
622 ;
623 ;
624 ;
625 ;
626 ;
627 ;
628 ;
629 ;
630 ;
631 ;
632 ;
633 ;
634 ;
635 ;
636 ;
637 ;
638 ;
639 ;
640 ;
641 ;
642 ;
643 ;
644 ;
645 ;
646 ;
647 ;
648 ;
649 ;
650 ;
651 ;
652 ;
653 ;
654 ;
655 ;
656 ;
657 ;
658 ;
659 ;
660 ;
661 ;
662 ;
663 ;
664 ;
665 ;
666 ;
667 ;
668 ;
669 ;
670 ;
671 ;
672 ;
673 ;
674 ;
675 ;
676 ;
677 ;
678 ;
679 ;
680 ;
681 ;
682 ;
683 ;
684 ;
685 ;
686 ;
687 ;
688 ;
689 ;
690 ;
691 ;
692 ;
693 ;
694 ;
695 ;
696 ;
697 ;
698 ;
699 ;
700 ;
701 ;
702 ;
703 ;
704 ;
705 ;
706 ;
707 ;
708 ;
709 ;
710 ;
711 ;
712 ;
713 ;
714 ;
715 ;
716 ;
717 ;
718 ;
719 ;
720 ;
721 ;
722 ;
723 ;
724 ;
725 ;
726 ;
727 ;
728 ;
729 ;
730 ;
731 ;
732 ;
733 ;
734 ;
735 ;
736 ;
737 ;
738 ;
739 ;
740 ;
741 ;
742 ;
743 ;
744 ;
745 ;
746 ;
747 ;
748 ;
749 ;
750 ;
751 ;
752 ;
753 ;
754 ;
755 ;
756 ;
757 ;
758 ;
759 ;
760 ;
761 ;
762 ;
763 ;
764 ;
765 ;
766 ;
767 ;
768 ;
769 ;
770 ;
771 ;
772 ;
773 ;
774 ;
775 ;
776 ;
777 ;
778 ;
779 ;
780 ;
781 ;
782 ;
783 ;
784 ;
785 ;
786 ;
787 ;
788 ;
789 ;
790 ;
791 ;
792 ;
793 ;
794 ;
795 ;
796 ;
797 ;
798 ;
799 ;
800 ;
801 ;
802 ;
803 ;
804 ;
805 ;
806 ;
807 ;
808 ;
809 ;
810 ;
811 ;
812 ;
813 ;
814 ;
815 ;
816 ;
817 ;
818 ;
819 ;
820 ;
821 ;
822 ;
823 ;
824 ;
825 ;
826 ;
827 ;
828 ;
829 ;
830 ;
831 ;
832 ;
833 ;
834 ;
835 ;
836 ;
837 ;
838 ;
839 ;
840 ;
841 ;
842 ;
843 ;
844 ;
845 ;
846 ;
847 ;
848 ;
849 ;
850 ;
851 ;
852 ;
853 ;
854 ;
855 ;
856 ;
857 ;
858 ;
859 ;
860 ;
861 ;
862 ;
863 ;
864 ;
865 ;
866 ;
867 ;
868 ;
869 ;
870 ;
871 ;
872 ;
873 ;
874 ;
875 ;
876 ;
877 ;
878 ;
879 ;
880 ;
881 ;
882 ;
883 ;
884 ;
885 ;
886 ;
887 ;
888 ;
889 ;
890 ;
891 ;
892 ;
893 ;
894 ;
895 ;
896 ;
897 ;
898 ;
899 ;
900 ;
901 ;
902 ;
903 ;
904 ;
905 ;
906 ;
907 ;
908 ;
909 ;
910 ;
911 ;
912 ;
913 ;
914 ;
915 ;
916 ;
917 ;
918 ;
919 ;
920 ;
921 ;
922 ;
923 ;
924 ;
925 ;
926 ;
927 ;
928 ;
929 ;
930 ;
931 ;
932 ;
933 ;
934 ;
935 ;
936 ;
937 ;
938 ;
939 ;
940 ;
941 ;
942 ;
943 ;
944 ;
945 ;
946 ;
947 ;
948 ;
949 ;
950 ;
951 ;
952 ;
953 ;
954 ;
955 ;
956 ;
957 ;
958 ;
959 ;
960 ;
961 ;
962 ;
963 ;
964 ;
965 ;
966 ;
967 ;
968 ;
969 ;
970 ;
971 ;
972 ;
973 ;
974 ;
975 ;
976 ;
977 ;
978 ;
979 ;
980 ;
981 ;
982 ;
983 ;
984 ;
985 ;
986 ;
987 ;
988 ;
989 ;
990 ;
991 ;
992 ;
993 ;
994 ;
995 ;
996 ;
997 ;
998 ;
999 ;
1000 ;

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/044,995
FILING DATE: 15-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: <Unknown>
APPLICATION NUMBER: US 07/920,286
FILING DATE: 14-OCT-1992
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-044-995-23
Query Match 100.0%; Score 20; DB 15; Length 2894;
Best Local Similarity 100.0%; Pred. No. 1.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCERLASCRPLTDFDQGWG 20
Db 451 GCERLASCRPLTDFDQGWG 470
RESULT 12
US-09-916-359-2
Sequence 2; Application US/09916359
Patent No. US20020034734A1
GENERAL INFORMATION:
APPLICANT: Veronique Barban
TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
TREATING C HEPATITIS
FILE REFERENCE: PMCF97-03A
CURRENT APPLICATION NUMBER: US/09/916,359
CURRENT FILING DATE: 2001-07-26
PRIOR APPLICATION NUMBER: 09/388,874
PRIOR FILING DATE: 1999-09-02
PRIOR APPLICATION NUMBER: 97/02,887
PRIOR FILING DATE: 1997-03-06
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 3011
TYPE: PRT
ORGANISM: Virus
US-09-916-359-2
Query Match 100.0%; Score 20; DB 9; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.6e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCERLASCRPLTDFDQGWG 20
Db 451 GCERLASCRPLTDFDQGWG 470

Db 451 GCPERLASCRPLTDFDQGWG 470

RESULT 13

US-10-232-643-6

; Sequence 6, Application US/10232643

; Publication No. US20030129586A1

GENERAL INFORMATION:

APPLICANT: HOUGHTON, MICHAEL

CHOO, QUI-LIM

HAN, JANG

CHOE, JOONHO

TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING

HELICASE ACTIVITY AND IMPROVED SOLUBILITY

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: CHIRON CORPORATION

STREET: Intellectual Property - R440, P.O. Box 8097

CITY: Emeryville

STATE: California

COUNTRY: U.S.A.

ZIP: 94662-8097

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/232,643

FILING DATE: 30-Aug-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/483,799

FILING DATE: <Unknown>

APPLICATION NUMBER: 08/529,169

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Harbin, Alisa A.

REGISTRATION NUMBER: 33,895

REFERENCE/DOCKET NUMBER: 0100.005

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 923-3274

TELEFAX: (510) 655-3542

TELEX: n/a

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 3011 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Duplication

LOCATION: 9

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Lys or Arg"

FEATURE:

NAME/KEY: Duplication

LOCATION: 11

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Asn or Thr"

FEATURE:

NAME/KEY: Duplication

LOCATION: 176

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Ile or Thr"

FEATURE:

NAME/KEY: Duplication

LOCATION: 334

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Met or Val"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2502

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Leu or Phe"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1454

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Cys or Tyr"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1471

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Thr or Ser"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1877

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Glu or Gly"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1948

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Leu or His"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1949

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Ser or Cys"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2021

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Gly or Val"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2349

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Thr or Ser"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2385

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Tyr or Phe"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2386

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Ser or

Ala"

FEATURE:

NAME/KEY: Duplication

LOCATION: 603

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Leu or Ile"

FEATURE:

NAME/KEY: Duplication

LOCATION: 848

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Tyr or Asn"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1114

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Pro or Ser"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1117

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Ser or Thr"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1276

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Pro or Leu"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1454

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Cys or Tyr"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1471

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Thr or Ser"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1877

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Glu or Gly"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1948

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Leu or His"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1949

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Ser or Cys"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2021

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Gly or Val"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2349

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Thr or Ser"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2385

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Tyr or Phe"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2386

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Ser or

Ala"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2502

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Leu or Phe"

```
;
;
; FEATURE:
;   NAME/KEY: Duplication
;   LOCATION: 2690
;   OTHER INFORMATION: /note= "There exists a
;   heterogeneity at this position - Xaa = Arg or Gly"
;
; FEATURE:
;   NAME/KEY: Duplication
;   LOCATION: 2921
;   OTHER INFORMATION: /note= "There exists a
;   heterogeneity at this position - Xaa = Arg or Gly"
;
; FEATURE:
;   NAME/KEY: Duplication
;   LOCATION: 2996
;   OTHER INFORMATION: /note= "There exists a
;   heterogeneity at this position - Xaa = Leu or Pro"
;
;   SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-232-643-6

Query Match      100.0%; Score 20; DB 16; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.6e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 GCPERLASCRPLTDFDQGWG 20
    |||||
Db 451 GCPERLASCRPLTDFDQGWG 470
```

```
RESULT 14
US-10-318-200-10
; Sequence 10, Application US/10318200
; Publication No. US20030129746A1
; GENERAL INFORMATION:
;   APPLICANT: MAERTENS, GEERT
;   APPLICANT: DELPE, ERIK
;   TITLE OF INVENTION: BUYSE, MARIE-ANGE
;   TITLE OF INVENTION: ANTIBODIES IN VIRAL ENVELOPE PROTEINS AND SPECIFIC
;   TITLE OF INVENTION: ANTIBODIES DIRECTED AGAINST THESE EPITOPES: USE FOR
;   TITLE OF INVENTION: DETECTION OF HCV VIRAL ANTIGEN IN HOST TISSUE
;   FILE REFERENCE: 2551-47
;   CURRENT APPLICATION NUMBER: US/10/318,200
;   CURRENT FILING DATE: 2002-12-13
;   PRIOR APPLICATION NUMBER: US/09/645,470
;   PRIOR FILING DATE: 2000-08-24
;   PRIOR APPLICATION NUMBER: PCT/EP99/02154
;   PRIOR FILING DATE: 1999-03-29
;   PRIOR APPLICATION NUMBER: EP 98870060.5
;   PRIOR FILING DATE: 1998-03-27
;   NUMBER OF SEQ ID NOS: 41
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 10
;   LENGTH: 38
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-10-318-200-10
```

```
Query Match      50.0%; Score 10; DB 16; Length 38;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 GCPERLASCR 10
    |||||
Db 22 GCPERLASCR 31
```

```
RESULT 15
US-09-921-397-80
; Sequence 80, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
;   APPLICANT: HYBRIGENICS
;   TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
;   TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
;   TITLE OF INVENTION: applications thereof
```

```
;
;
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 80
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-80

Query Match      50.0%; Score 10; DB 10; Length 91;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCR 10
    |||||
Db 72 GCPERLASCR 81

Search completed: November 21, 2003, 22:19:37
Job time : 19.5 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 163.1 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-76

Perfect score: 20

Sequence: 1 GCPERLASCRPLTDFDQGWG 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents AA Main:*

- 1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/1/paa/US097A_COMB.pep.*
- 22: /cgn2_6/ptodata/1/paa/US097B_COMB.pep.*
- 23: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
- 24: /cgn2_6/ptodata/1/paa/US099A_COMB.pep.*
- 25: /cgn2_6/ptodata/1/paa/US099B_COMB.pep.*
- 26: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
- 27: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
- 28: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
- 29: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
- 30: /cgn2_6/ptodata/1/paa/US104_COMB.pep.*
- 31: /cgn2_6/ptodata/1/paa/US106_COMB.pep.*
- 32: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	13	US-08-974-685-32
					Sequence 32, Appl

2	20	100.0	20	13	US-08-974-690-32	Sequence 32, Appl
3	20	100.0	20	13	US-08-974-690A-32	Sequence 32, Appl
4	20	100.0	20	13	US-08-974-690B-32	Sequence 32, Appl
5	20	100.0	20	13	US-08-974-690C-32	Sequence 32, Appl
6	20	100.0	20	23	US-09-899-303-76	Sequence 76, Appl
7	20	100.0	20	23	US-09-899-303A-76	Sequence 76, Appl
8	20	100.0	20	25	US-09-973-025-76	Sequence 76, Appl
9	20	100.0	20	25	US-09-973-025-76	Sequence 76, Appl
10	20	100.0	20	25	US-09-995-808-76	Sequence 76, Appl
11	20	100.0	20	25	US-09-995-808-76	Sequence 76, Appl
12	20	100.0	20	26	US-10-020-510-76	Sequence 76, Appl
13	20	100.0	20	29	US-10-321-798-76	Sequence 76, Appl
14	20	100.0	37	9	US-08-507-740-18	Sequence 18, Appl
15	20	100.0	37	13	US-08-905-054-18	Sequence 18, Appl
16	20	100.0	37	13	US-08-905-054B-18	Sequence 18, Appl
17	20	100.0	76	8	US-08-403-590B-79	Sequence 79, Appl
18	20	100.0	76	8	US-08-444-112-79	Sequence 79, Appl
19	20	100.0	122	8	US-08-436-986-5	Sequence 5, Appl
20	20	100.0	254	1	PCT-US00-26395-3	Sequence 3, Appl
21	20	100.0	254	18	US-09-407-430-3	Sequence 3, Appl
22	20	100.0	278	8	US-08-471-498-15	Sequence 15, Appl
23	20	100.0	333	30	US-10-445-724-3	Sequence 3, Appl
24	20	100.0	333	32	US-60-409-909-4	Sequence 4, Appl
25	20	100.0	363	1	PCT-US00-26395-2	Sequence 2, Appl
26	20	100.0	363	18	US-09-407-430-2	Sequence 2, Appl
27	20	100.0	416	8	US-08-436-966-9	Sequence 9, Appl
28	20	100.0	454	8	US-08-403-590B-73	Sequence 73, Appl
29	20	100.0	454	8	US-08-444-112-73	Sequence 73, Appl
30	20	100.0	463	3	US-07-748-561-1	Sequence 1, Appl
31	20	100.0	463	3	US-07-989-843-31	Sequence 31, Appl
32	20	100.0	463	8	US-08-463-849-31	Sequence 31, Appl
33	20	100.0	463	8	US-08-463-884-31	Sequence 31, Appl
34	20	100.0	463	20	US-09-690-359-31	Sequence 31, Appl
35	20	100.0	480	8	US-08-471-498-14	Sequence 14, Appl
36	20	100.0	531	21	US-09-721-480-7	Sequence 7, Appl
37	20	100.0	622	3	US-07-748-561-4	Sequence 4, Appl
38	20	100.0	622	3	US-07-989-843-34	Sequence 34, Appl
39	20	100.0	622	8	US-08-463-849-34	Sequence 34, Appl
40	20	100.0	622	8	US-08-463-884-34	Sequence 34, Appl
41	20	100.0	622	20	US-09-690-359-34	Sequence 34, Appl
42	20	100.0	637	1	PCT-US02-20676-4	Sequence 4, Appl
43	20	100.0	637	1	PCT-US02-30423-2	Sequence 2, Appl
44	20	100.0	637	1	PCT-US02-31486-2	Sequence 2, Appl
45	20	100.0	637	27	US-10-187-257-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-08-974-685-32

Sequence 32, Application US/08974685

GENERAL INFORMATION:

APPLICANT: LEROUIX-ROELS, GEERT

DELEYS, ROBERT

MAERTENS, GEERT

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF

HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 181

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHVE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,685
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-08-974-685-32

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 GCPERLASCRPLTDFDQGWG 20

RESULT 2

US-08-974-690-32
Sequence 32, Application US/08974690
GENERAL INFORMATION:
APPLICANT: LEROUIX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 166
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P. C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/635,886
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

US-08-974-690-32

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 GCPERLASCRPLTDFDQGWG 20

RESULT 3

US-08-974-690A-32
Sequence 32, Application US/08974690A
GENERAL INFORMATION:
APPLICANT: LEROUIX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P. C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690A
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-08-974-690A-32

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 GCPERLASCRPLTDFDQGWG 20

RESULT 4

US-08-974-690B-32
Sequence 32, Application US/08974690B
GENERAL INFORMATION:
APPLICANT: LEROUIX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHUYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690B
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-08-974-690B-32

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCPLASCRPLTDFDQGWG 20
|||||

DB 1 GCPLASCRPLTDFDQGWG 20
|||||

RESULT 5
US-08-974-690C-32
; Sequence 32, Application US/08974690C
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/08/974,690C
; CURRENT FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 32
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-974-690C-32

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCPLASCRPLTDFDQGWG 20

DB 1 GCPLASCRPLTDFDQGWG 20
|||||

RESULT 6
US-09-899-303-76
; Sequence 76, Application US/09899303
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 76:
US-09-899-303-76

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCPLASCRPLTDFDQGWG 20
|||||

DB 1 GCPLASCRPLTDFDQGWG 20
|||||

RESULT 7
US-09-899-303A-76
; Sequence 76, Application US/09899303A
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.


```
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCPERLASCRPLTDFDQGWG 20
Db 1 GCPERLASCRPLTDFDQGWG 20

RESULT 11
US-09-995-860-76
; Sequence 76, Application US/09995860
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 76
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-76

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCPERLASCRPLTDFDQGWG 20
Db 1 GCPERLASCRPLTDFDQGWG 20

RESULT 12
US-10-020-510-76
; Sequence 76, Application US/10020510
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-72
; CURRENT APPLICATION NUMBER: US/10/020,510
; CURRENT FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 76
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-020-510-76

Query Match 100.0%; Score 20; DB 26; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCPERLASCRPLTDFDQGWG 20
Db 1 GCPERLASCRPLTDFDQGWG 20

RESULT 13
US-10-321-798-76
; Sequence 76, Application US/10321798
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-93
; CURRENT APPLICATION NUMBER: US/10/321,798
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: 60/418,358
; PRIOR FILING DATE: 2002-10-16
```

```
; PRIOR APPLICATION NUMBER: 10/020,510
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 76
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-321-798-76

Query Match 100.0%; Score 20; DB 29; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCPERLASCRPLTDFDQGWG 20
Db 1 GCPERLASCRPLTDFDQGWG 20

RESULT 14
US-08-507-740-18
; Sequence 18, Application US/08507740
; GENERAL INFORMATION:
; APPLICANT: LESNIEWSKI, RICHARD R.
; APPLICANT: LEUNG, TAT K.
; TITLE OF INVENTION: HEPATITIS C ASSAY
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES CHAD377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: U.S.A.
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/507,740
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/373,920
; FILING DATE:
; APPLICATION NUMBER: US/08/183,207
; FILING DATE:
; APPLICATION NUMBER: US 07/760,292
; FILING DATE: 16-SEP-1991
; APPLICATION NUMBER: US 07/456,162
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/610,180
; FILING DATE: 07-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMBSKIP, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4767-US.P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 37 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-507-740-18

Query Match 100.0%; Score 20; DB 9; Length 37;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
```

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
| | | | | | | | | | | | | | | | | | | | | |
Db 5 GCPERLASCRPLTDFDQGWG 24

RESULT 15
US-08-905-054-18
; Sequence 18, Application US/08905054
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Lesniewski, Richard R.
; APPLICANT: Leung, Tat K.
; TITLE OF INVENTION: HEPATITIS C ASSAY
; FILE REFERENCE: 4767.US.C7
; CURRENT APPLICATION NUMBER: US/08/905,054
; CURRENT FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/707,355
; PRIOR FILING DATE: 1996-09-04
; PRIOR APPLICATION NUMBER: US 08/507,740
; PRIOR FILING DATE: 1995-07-26
; PRIOR APPLICATION NUMBER: US 08/373,920
; PRIOR FILING DATE: 1995-01-17
; PRIOR APPLICATION NUMBER: US 08/183,207
; PRIOR FILING DATE: 1994-01-18
; PRIOR APPLICATION NUMBER: US 07/760,292
; PRIOR FILING DATE: 1991-09-16
; PRIOR APPLICATION NUMBER: US 07/610,180
; PRIOR FILING DATE: 1990-11-07
; PRIOR APPLICATION NUMBER: US 07/456,162
; PRIOR FILING DATE: 1989-12-22
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-08-905-054-18

Query Match 100.0%; Score 20; DB 13; Length 37;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
| | | | | | | | | | | | | | | | | | | | | |
Db 5 GCPERLASCRPLTDFDQGWG 24

Search completed: November 21, 2003, 22:09:51
Job time : 164.1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.55' Seconds
(without alignments)
117.536 Million cell updates/sec

Title: US-09-973-025-76

Perfect score: 20

Sequence: 1 GCPERLASCRPLTDFDQGWG 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA_New.*

- 1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	6	US-10-651-165-32
2	20	100.0	333	1	PCT-US03-19834-3
3	20	100.0	333	1	US-10-655-562-4
4	20	100.0	637	1	PCT-US03-33610-4
5	20	100.0	3011	1	PCT-US03-19834-2
6	10	50.0	30	6	US-10-296-734-466
7	10	50.0	30	6	US-10-296-734-468
8	10	50.0	38	6	US-10-685-435-19
9	10	50.0	1997	6	US-10-296-734-816
10	10	50.0	3011	6	US-10-296-734-406
11	10	50.0	5985	6	US-10-296-734-810
12	8	40.0	20	6	US-10-651-165-31
13	8	40.0	20	6	US-10-651-165-33
14	7	35.0	30	6	US-10-296-734-470
15	7	35.0	2010	6	US-10-296-734-814
16	6	30.0	106	6	US-10-425-114A-47745
17	6	30.0	118	6	US-10-425-114A-58262
18	6	30.0	176	5	US-09-638-693-158
19	6	30.0	176	5	US-09-873-224A-158
20	6	30.0	20	7	US-60-495-114-1554
21	6	30.0	299	6	US-10-412-699B-600
22	6	30.0	301	6	US-10-425-114A-56112
23	6	30.0	336	6	US-10-679-063-2311
24	6	30.0	370	6	US-10-679-063-16458
25	6	30.0	390	6	US-10-679-063-2533
26	6	30.0	390	6	US-10-679-063-2534

27	6	30.0	459	6	US-10-425-114A-55352	Sequence 55352, A
28	6	30.0	514	6	US-10-679-063-15534	Sequence 15534, A
29	6	30.0	769	6	US-10-412-699B-1938	Sequence 1938, Ap
30	6	30.0	771	6	US-10-679-063-2598	Sequence 2598, Ap
31	6	30.0	832	6	US-10-425-114A-72876	Sequence 72876, A
32	6	30.0	1026	1	PCT-US03-20409-3	Sequence 3, Appli
33	6	30.0	1814	6	US-10-367-094-162	Sequence 162, App
34	6	30.0	5317	1	PCT-US03-29834-59	Sequence 59, Appl
35	6	30.0	5317	6	US-10-668-767-59	Sequence 59, Appl
36	5	25.0	38	6	US-10-684-496-165	Sequence 165, App
37	5	25.0	78	5	US-09-897-516A-5240	Sequence 5240, Ap
38	5	25.0	83	6	US-10-425-114A-40004	Sequence 40004, A
39	5	25.0	89	6	US-10-425-114A-37670	Sequence 37670, A
40	5	25.0	91	6	US-10-473-305-896	Sequence 896, App
41	5	25.0	95	6	US-10-425-114A-41244	Sequence 41244, A
42	5	25.0	105	6	US-10-461-060-14	Sequence 14, Appl
43	5	25.0	107	5	US-09-897-516A-7462	Sequence 7462, Ap
44	5	25.0	108	6	US-10-311-129-29	Sequence 29, Appl
45	5	25.0	113	5	US-09-638-693-160	Sequence 160, App

ALIGNMENTS

RESULT 1
US-10-651-165-32
; Sequence 32, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651.165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974.690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-32

Query Match 100.0%; Score 20; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
|||||
Db 1 GCPERLASCRPLTDFDQGWG 20

RESULT 2
PCT-US03-19834-3
; Sequence 3, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; TITLE OF INVENTION: METHODS FOR TREATING HYPERLIPIDEMIA
; FILE REFERENCE: IOWA:045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158

```
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-3

Query Match      100.0%; Score 20; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 3.4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GCPERLASCRPLTDFDQGWG 20
Db      69 GCPERLASCRPLTDFDQGWG 88

RESULT 3
US-10-655-562-4
; Sequence 4, Application US/10655562
; GENERAL INFORMATION:
; APPLICANT: BISHOP-HURLEY, SHARON L.
; APPLICANT: SCHMIDT, FRANCIS J.
; APPLICANT: SMITH, ARNOLD L.
; TITLE OF INVENTION: PRAGE-DISPLAY PEPTIDES AS NOVEL ANTIMICROBIAL AGENTS
; FILE REFERENCE: UVM0:022US
; CURRENT APPLICATION NUMBER: US/10/655.562
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: 60,409,909
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-10-655-562-4

Query Match      100.0%; Score 20; DB 6; Length 333;
Best Local Similarity 100.0%; Pred. No. 3.4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GCPERLASCRPLTDFDQGWG 20
Db      69 GCPERLASCRPLTDFDQGWG 88

RESULT 4
PCT-US03-33610-4
; Sequence 4, Application PC/TUS0333610
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: SELBY, Mark
; APPLICANT: PALIARD, Xavier
; TITLE OF INVENTION: ACTIVATION OF HCV-SPECIFIC T-CELLS
; FILE REFERENCE: 2300-1612.60
; CURRENT APPLICATION NUMBER: PCT/US03/33610
; CURRENT FILING DATE: 2003-10-24
; PRIOR APPLICATION NUMBER: 10/281,341
; PRIOR FILING DATE: 2002-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-10-655-562-4

Query Match      100.0%; Score 20; DB 1; Length 637;
Best Local Similarity 100.0%; Pred. No. 5.3e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GCPERLASCRPLTDFDQGWG 20
Db      279 GCPERLASCRPLTDFDQGWG 298

RESULT 5
PCT-US03-19834-2
; Sequence 2, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; FILE REFERENCE: IOWA-045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
PCT-US03-19834-2

Query Match      100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.5e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GCPERLASCRPLTDFDQGWG 20
Db      451 GCPERLASCRPLTDFDQGWG 470

RESULT 6
US-10-296-734-466
; Sequence 466, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 466
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 30
US-10-296-734-466

Query Match      50.0%; Score 10; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GCPERLASCR 10
Db      18 GCPERLASCR 27
```

RESULT 7
US-10-296-734-468
; Sequence 468, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 468
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 31
US-10-296-734-468

Query Match 50.0%; Score 10; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCR 10
Db 3 GCPERLASCR 12

RESULT 8
US-10-685-435-19
; Sequence 19, Application US/10685435
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Multi-mer peptides derived from Hepatitis C Virus
; TITLE OF INVENTION: envelope proteins for diagnostic use and vaccination
; FILE REFERENCE: PCT98.75.HCV30
; CURRENT APPLICATION NUMBER: US/10/685,435
; CURRENT FILING DATE: 2003-10-16
; PRIOR APPLICATION NUMBER: US/09/566,266B
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 97870179.5
; PRIOR FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: patenting Ver. 2.1
; SEQ ID NO 19
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-685-435-19

Query Match 50.0%; Score 10; DB 6; Length 38;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCR 10
Db 22 GCPERLASCR 31

RESULT 9
US-10-296-734-816
; Sequence 816, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine

; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 816
; LENGTH: 1997
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette C
US-10-296-734-816

Query Match 50.0%; Score 10; DB 6; Length 1997;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCR 10
Db 618 GCPERLASCR 627

RESULT 10
US-10-296-734-406
; Sequence 406, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a consensus polyprotein
US-10-296-734-406

Query Match 50.0%; Score 10; DB 6; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.046;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCR 10
Db 451 GCPERLASCR 460

RESULT 11
US-10-296-734-810
; Sequence 810, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 810
; LENGTH: 5985
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:

```
; OTHER INFORMATION: HepC 1a savine
US-10-296-734-810

Query Match          50.0%; Score 10; DB 6; Length 5985;
Best Local Similarity 100.0%; Pred. No. 0.074;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCR 10
Db 4638 GCPERLASCR 4647

RESULT 12
US-10-651-165-31
; Sequence 31, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-31

Query Match          40.0%; Score 8; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLAS 8
Db 13 GCPERLAS 20

RESULT 13
US-10-651-165-33
; Sequence 33, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-33

; OTHER INFORMATION: HepC 1a savine
US-10-296-734-810

Query Match          40.0%; Score 8; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 TDFDQGWG 20
Db 1 TDFDQGWG 8

RESULT 14
US-10-296-734-470
; Sequence 470, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 470
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 32
US-10-296-734-470

Query Match          35.0%; Score 7; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 DFDQGWG 20
Db 1 DFDQGWG 7

RESULT 15
US-10-296-734-814
; Sequence 814, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 814
; LENGTH: 2010
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette B
US-10-296-734-814

Query Match          35.0%; Score 7; DB 6; Length 2010;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 DFDQGWG 20
Db 1951 DFDQGWG 1957

Search completed: November 21, 2003, 22:12:54
Job time : 8.55 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:46:27 ; Search time 9.25 Seconds
(without alignments)
207.932 Million cell updates/sec

Title: US-09-973-025-76

Perfect score: 20

Sequence: 1 GCPERLASCRPLTDFDQGWG 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	138	2 S24081	envelope protein -
2	20	100.0	513	2 A44150	structural protein
3	20	100.0	3011	1 GNWVC3	genome polyprotein
4	15	75.0	138	2 S24080	envelope protein -
5	11	55.0	138	2 S24106	envelope protein -
6	10	50.0	235	2 S32747	genome polyprotein
7	10	50.0	3011	1 GNWVCH	genome polyprotein
8	10	50.0	3011	1 S40770	genome polyprotein
9	7	35.0	138	2 S24075	envelope protein -
10	7	35.0	138	2 S24074	envelope protein -
11	7	35.0	577	2 AG0501	thrombomodulin pre
12	7	35.0	640	2 JQ1584	genome polyprotein
13	7	35.0	716	2 JQ1366	polyprotein - hepa
14	6	30.0	112	2 A66604	periplasmic divale
15	6	30.0	112	2 B72019	periplasmic divale
16	6	30.0	127	2 B48776	polyprotein (E2/NS
17	6	30.0	130	2 G70813	probable lpqs prot
18	6	30.0	133	2 C48776	polyprotein (E2/NS
19	6	30.0	133	2 D48776	polyprotein (E2/NS
20	6	30.0	133	2 E48776	polyprotein (E2/NS
21	6	30.0	138	2 S24097	envelope protein -
22	6	30.0	138	2 S24096	envelope protein -
23	6	30.0	138	2 S24078	envelope protein -
24	6	30.0	138	2 S24092	envelope protein -
25	6	30.0	138	2 S24099	envelope protein -
26	6	30.0	138	2 S24073	envelope protein -
27	6	30.0	138	2 S24105	envelope protein -
28	6	30.0	138	2 S24094	envelope protein -
29	6	30.0	138	2 S24068	envelope protein -

30 6 30.0 138 2 S24069 envelope protein -
31 6 30.0 149 1 LYBPD1 lytic enzyme (EC 3
32 6 30.0 156 2 F85061 hypothetical prote
33 6 30.0 215 2 E82958 conserved hypother
34 6 30.0 227 2 T21783 hypothetical prote
35 6 30.0 249 2 T40405 hypothetical prote
36 6 30.0 303 1 S27618 5-dehydro-4-deoxyg
37 6 30.0 315 2 P80164 envelope glycoprot
38 6 30.0 327 2 F96814 D-mannosate hydrol
39 6 30.0 361 2 G83782 hypothetical prote
40 6 30.0 364 2 H70880 hypothetical prote
41 6 30.0 370 1 S68187 choline-phosphate
42 6 30.0 400 2 H91163 probable DNA proce
43 6 30.0 431 2 A86010 probable DNA proce
44 6 30.0 441 2 E84264 isochorismate synt
45 6 30.0 449 2 T29425 4-carboxymuconolac

ALIGNMENTS

RESULT 1

S24081

envelope protein - hepatitis C virus (fragment)

N:Contains: envelope protein gp35; envelope protein gp70

C:Species: hepatitis C virus

C:Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000

C:Accession: S24081

R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraio, K.; Ohkoshi,

Virus Res. 22, 107-123, 1992

A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C vir

A:Reference number: S24068; MUID:92230394; PMID:1314471

A:Accession: S24081

A:Molecule type: genomic RNA

A:Residues: 1-138 <KAT>

A:Cross-references: EMBL:X60569

A:Experimental source: isolate RE43

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: envelope protein; glycoprotein; polyprotein

F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>

F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>

F:16-42/Region: hypervariable 1 #status predicted

F:106-112/Region: hypervariable 2 #status predicted

F:49,55,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 20; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCPERLASCRPLTDFDQGWG 20

DB 83 GCPERLASCRPLTDFDQGWG 102

RESULT 2

A44150

structural protein - hepatitis C virus

C:Species: hepatitis C virus

C:Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 17-Nov-2000

C:Accession: A44150

R:Ching, W.M.; Wychowski, C.; Beach, M.J.; Wang, H.; Davies, C.L.; Carl, M.; Bradley, D

Proc. Natl. Acad. Sci. U.S.A. 89, 3190-3194, 1992

A:Title: Interaction of immune sera with synthetic peptides corresponding to the structu

A:Reference number: A44150; MUID:92228749; PMID:1373489

A:Accession: A44150

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: genomic RNA

A:Residues: 1-513 <CHI>

C:Superfamily: hepatitis C virus genome polyprotein

Query Match 100.0%; Score 20; DB 2; Length 513;
Best Local Similarity 100.0%; Pred. No. 8.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
 |||||
 Db 451 GCPERLASCRPLTDFDQGWG 470

RESULT 3
 GNMVC3
 genome polypeptide - hepatitis C virus (strain HCV-1)
 N:Contains: capsid protein C; envelope protein M; Hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)
 C:Species: hepatitis C virus
 C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001
 C:Accession: A39166; PQ0403; PQ0404
 R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coito, Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
 A:Title: Genetic organization and diversity of the hepatitis C virus.
 A:Reference number: A39166; MUID:91172826; PMID:1848704
 A:Accession: A39166
 A:Molecule type: mRNA
 A:Residues: 1-3011 <CHO>
 A:Cross-references: GB:M62321; NID:G329873; PIDN:AAA45676.1; PID:G329874
 R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L. J. Gen. Virol. 73, 1131-1141, 1992
 A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to other hepatitis C virus strains.
 A:Reference number: PQ0393; MUID:92268871; PMID:1316939
 A:Accession: PQ0403
 A:Molecule type: genomic RNA
 A:Residues: 1577-1633 <CHA>
 A:Cross-references: DBJ:D10128
 A:Experimental source: isolates E-b16
 A:Accession: PQ0404
 A>Status: preliminary
 A:Molecule type: genomic RNA
 A:Residues: 1577-1633 <CH2>
 A:Experimental source: isolates E-b17
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein
 F:11-15/Product: capsid protein C #status predicted <CPC>
 F:116-19/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-725/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196, 209, 234, 305, 325, 417, 423, 430, 448, 476, 532, 540, 556, 576, 623, 645, 1213, 1255, 2041, 2077, 22

Query Match 100.0%; Score 20; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 3.6e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
 |||||
 Db 451 GCPERLASCRPLTDFDQGWG 470

RESULT 4
 S24080
 envelope protein - hepatitis C virus (fragment)
 N:Contains: envelope protein gp35; envelope protein gp70
 C:Species: hepatitis C virus
 C:Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
 C:Accession: S24080
 R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraishi, K.; Ohkoshi, Virus Res. 22, 107-123, 1992
 A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
 A:Reference number: S24068; MUID:92230394; PMID:1314471
 A:Accession: S24080

A:Molecule type: genomic RNA
 A:Residues: 1-138 <KAT>
 A:Cross-references: EMBL:X60568
 A:Experimental source: isolate RE42
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: envelope protein; glycoprotein; polyprotein
 F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
 F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
 F:16-42/Region: hypervariable 1 #status predicted
 F:106-112/Region: hypervariable 2 #status predicted
 F:106, 55, 62, 80/Binding site: carbohydrate (Asn) (covalent) #status predicted
 Query Match 75.0%; Score 15; DB 2; Length 138;
 Best Local Similarity 100.0%; Pred. No. 5.7e-09;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDF 15
 |||||
 Db 83 GCPERLASCRPLTDF 97

RESULT 5
 S24106
 envelope protein - hepatitis C virus (fragment)
 N:Contains: envelope protein gp35; envelope protein gp70
 C:Species: hepatitis C virus
 C:Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
 C:Accession: S24106
 R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraishi, K.; Ohkoshi, Virus Res. 22, 107-123, 1992
 A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
 A:Reference number: S24068; MUID:92230394; PMID:1314471
 A:Accession: S24106
 A:Molecule type: genomic RNA
 A:Residues: 1-138 <KAT>
 A:Cross-references: EMBL:X60589
 A:Experimental source: isolate RE75
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: envelope protein; glycoprotein; polyprotein
 F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
 F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
 F:16-42/Region: hypervariable 1 #status predicted
 F:106-112/Region: hypervariable 2 #status predicted
 F:106, 55, 62, 80/Binding site: carbohydrate (Asn) (covalent) #status predicted
 Query Match 55.0%; Score 11; DB 2; Length 138;
 Best Local Similarity 100.0%; Pred. No. 0.0001;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRP 11
 |||||
 Db 83 GCPERLASCRP 93

RESULT 6
 S32747
 genome polypeptide - hepatitis C virus (isolate HU-1) (fragment)
 N:Contains: envelope protein E2
 C:Species: hepatitis C virus
 A:Variety: isolate HU-1
 C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S32747
 R:Roggeandorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E. submitted to the EMBL Data Library, February 1993
 A:Description: Variability of the envelope regions of HCV in European isolates and its significance for the classification of HCV
 A:Reference number: S32741
 A:Accession: S32747
 A:Molecule type: genomic RNA
 A:Residues: 1-235 <ROG>
 A:Cross-references: EMBL:X72977; NID:9296112; PIDN:CAA51482.1; PID:G296113
 A:Experimental source: isolate HU-1
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: envelope protein; polyprotein

F:1-235/Product: envelope protein E2 #status predicted <MAT>

Query Match 50.0%; Score 10; DB 2; Length 235;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCR 10
|||||
Db 83 GCPERLASCR 92

RESULT 7

GNWCH

genome polyprotein - hepatitis C virus (strain H)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A36814; A41546
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: compared to GenBank, July 1992
A:Reference number: A36814

A:Accession: A36814

A:Molecule type: Genomic RNA

A:Residues: 1-3011 <INC>

A:Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738

R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.

Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991

A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: compared to GenBank, July 1992

A:Reference number: A41546; MUID:92052256; PMID:1658800

A:Contents: annotation

A:Note: neither amino acid nor nucleotide sequence is given

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

F:1-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitis C virus genome polyprotein

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match 50.0%; Score 10; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCR 10
|||||
Db 451 GCPERLASCR 460

RESULT 8

S40770

genome polyprotein - hepatitis C virus

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H)

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001

C:Accession: S40770; FC1285

R:Okamoto, H.

submitted to the EMBL Data Library, March 1992

A:Reference number: S40770

A:Accession: S40770

A:Molecule type: genomic RNA

A:Residues: 1-3011 <OKA>

A:Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tada, Jpn. J. Exp. Med. 60, 167-177, 1990

A:Title: The 5'-terminal sequence of the hepatitis C virus genome.

A:Reference number: FC1284; MUID:91013116; PMID:2170712

A:Accession: FC1285

A:Molecule type: genomic RNA

A:Residues: 1-513 <OK2>

A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512

A:Experimental source: isolate HC-J1

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine

F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitis C virus genome polyprotein

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 50.0%; Score 10; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCR 10
|||||
Db 451 GCPERLASCR 460

RESULT 9

S24075

envelope protein - hepatitis C virus (fragment)

N:Contains: envelope protein gp35; envelope protein gp70

C:Species: hepatitis C virus

C:Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000

C:Accession: S24075

R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraishi, K.; Ohkoshi,

Virus Res. 22, 107-123, 1992

A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C vir

A:Reference number: S24068; MUID:92230394; PMID:1314471

A:Accession: S24075

A:Molecule type: genomic RNA

A:Residues: 1-138 <KAT>

A:Cross-references: EMBL:X60564

A:Experimental source: isolate RE37B

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: envelope protein; glycoprotein; polyprotein

F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>

F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>

F:106-112/Region: hypervariable 1 #status predicted

F:106-112/Region: hypervariable 2 #status predicted

F:49,55,62,80,108/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 35.0%; Score 7; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ASCRPLT 13
|||||
Db 89 ASCRPLT 95

RESULT 10

S24074

envelope protein - hepatitis C virus (fragment)

N:Contains: envelope protein gp35; envelope protein gp70

C:Species: hepatitis C virus

C:Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000

C;Accession: S24074
R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, K.; Ohkoshi,
Virus Res. 22, 107-123, 1992
A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A;Reference number: S24068; MUID:92230394; PMID:1314471
A;Accession: S24074
A;Molecule type: genomic RNA
A;Residues: 1-138 <KAT>
A;Cross-references: ENBL:X60563
A;Experimental source: isolate RE37A
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: envelope protein; glycoprotein; polyprotein
F;1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F;16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F;16-42/Region: hypervariable 1 #status predicted
F;106-112/Region: hypervariable 2 #status predicted
F;49,55,62,80,108/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 35.0%; Score 7; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ASRCPLT 13
Db 89 ASRCPLT 95

RESULT 11
A60501
thrombomodulin precursor - mouse
N;Alternate names: fetomodulin
C;Species: Mus musculus (house mouse)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 16-Jul-1999
A;Accession: S08488; A32001; A60501
R;Dittman, W.A.; Majerus, P.W.
Nucleic Acids Res. 17, 802, 1989
A;Title: Sequence of a cDNA for mouse thrombomodulin and comparison of the predicted mou
A;Reference number: S08488; MUID:89128454; PMID:2536925
A;Accession: S08488
A;Molecule type: mRNA
A;Residues: 1-577 <DIT>
A;Cross-references: ENBL:X14432; NID:g54781; PIDN:CAA32597.1; PID:g54782
R;Dittman, W.A.; Kumada, T.; Sadler, J.E.; Majerus, P.W.
J. Biol. Chem. 263, 15815-15822, 1988
A;Title: The structure and function of mouse thrombomodulin. Phorbol myristate acetate s
A;Reference number: A32001; MUID:89008498; PMID:2844823
A;Accession: A32001
A;Molecule type: mRNA
A;Residues: 97-577 <DI2>
A;Cross-references: GB:J04060
R;Imada, S.; Yamaguchi, H.; Nagumo, M.; Katayanagi, S.; Iwasaki, H.; Imada, M.
Dev. Biol. 140, 113-122, 1990
A;Title: Identification of fetomodulin, a surface marker protein of fetal development, a
A;Reference number: A60501; MUID:90292331; PMID:2162790
A;Accession: A60501
A;Molecule type: protein
A;Residues: 19-22;330-343;479-489;545-555;562-575 <IMA>
C;Comment: Thrombomodulin binds to and internalizes with thrombin. It is also a cofactor
C;Superfamily: thrombomodulin; C-type lectin homology; EGF homology
C;Keywords: anticoagulant; membrane protein; phosphoprotein; receptor
F;24-165/Domain: C-type lectin homology <LCH>
F;244-279/Domain: EGF homology <EG1>
F;287-322/Domain: EGF homology <EG2>
F;328-361/Domain: EGF homology <EG3>
F;368-403/Domain: EGF homology <EG4>
F;407-438/Domain: EGF homology <EG5>
F;444-479/Domain: EGF homology <EG6>

Query Match 35.0%; Score 7; DB 2; Length 577;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ASRCPLT 13

Db 171 ASRCPLT 177

RESULT 12
JQ1584
genome polyprotein - hepatitis C virus (strain U.K.) (fragment)
N;Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural prot.
C;Species: hepatitis C virus
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
C;Accession: JQ1584
R;Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.
J. Gen. Virol. 73, 1521-1525, 1992
A;Title: Cloning and sequencing of the structural region and expression of putative core
A;Reference number: JQ1584; MUID:92300349; PMID:1318944
A;Accession: JQ1584
A;Molecule type: genomic RNA
A;Residues: 1-640 <KUM>
A;Cross-references: GB:X84079; NID:G643119; PIDN:CAA58888.1; PID:G643120
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; polypro
F;1-191/Product: core protein C #status predicted <CPC>
F;192-389/Product: envelope protein E1 #status predicted <EE1>
F;390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predicted <
F;196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn) (cova
Query Match 35.0%; Score 7; DB 2; Length 640;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 DFDQGWG 20
Db 464 DFDQGWG 470

RESULT 13
JQ1366
polyprotein - hepatitis C virus (French isolate) (fragments)
C;Species: hepatitis C virus
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
R;Kremsdorf, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.
J. Gen. Virol. 72, 2557-2561, 1991
A;Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication
A;Reference number: JQ1366; MUID:92013977; PMID:1655961
A;Accession: JQ1366
A;Molecule type: genomic RNA
A;Residues: 1-716 <KRE>
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: glycoprotein; polyprotein
F;84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #stat
Query Match 35.0%; Score 7; DB 2; Length 716;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 DFDQGWG 20
Db 131 DFDQGWG 137

RESULT 14
A86604
periplasmic divalent cation tolerance protein [imported] - Chlamydomophila pneumoniae (str
C;Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C;Accession: A86604
R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
Nucleic Acids Res. 28, 2311-2314, 2000
A;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A;Reference number: A86491; MUID:20330349; PMID:10871362
A;Accession: A86604
A;Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-112 <STO>
 A:Cross-references: GB:BA000008; NID:g8979281; PIDN:BAA99115.1; GSPDB:GNO0142
 A:Experimental source: strain J138
 C:Genetics:
 A:Gene: cutA
 C:Superfamily: divalent cation tolerance protein cutA1

Query Match 30.0%; Score 6; DB 2; Length 112;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ERLASC 9
 |||||
 DB 26 ERLASC 31

RESULT 15

B72019
 periplasmic divalent cation tolerance protein CutA1 CP0959 [imported] - Chlamydomophila pneumoniae
 C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 02-Sep-2000
 C:Accession: B72019; F81520
 R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Ölinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999
 A:Title: Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:99206606; PMID:10192388
 A:Accession: B72019
 A:Molecule type: DNA
 A:Residues: 1-112 <ARN>
 A:Cross-references: GB:AE001671; GB:AE001363; NID:g4377226; PIDN:AAI9045.1; PID:g437722
 A:Experimental source: strain CML029
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, J. C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
 A:Reference number: AB1500; MUID:20150255; PMID:10684935
 A:Accession: F81520
 A:Molecule type: DNA
 A:Residues: 1-112 <REA>
 A:Cross-references: GB:AE002253; GB:AE002161; NID:g7189859; PIDN:AAF38739.1; PID:g718987
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: cutA; CP0959
 C:Superfamily: divalent cation tolerance protein cutA1

Query Match 30.0%; Score 6; DB 2; Length 112;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ERLASC 9
 |||||
 DB 26 ERLASC 31

Search completed: November 21, 2003, 21:11:30
 Job time : 9.25 secs .

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:52 ; Search time 4.85 Seconds
(without alignments)
193.925 Million cell updates/sec

Title: US-09-973-025-76

Perfect score: 20

Sequence: 1 GCPERLASCRPLTDFDQGWG 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	3011	1 POLG_HCV1	P26664 h genome po
2	10	50.0	513	1 POLG_HCVJ2	P27959 hepatitis c
3	10	50.0	3011	1 POLG_HCVH	P27958 h genome po
4	7	35.0	577	1 TRBM_MOUSE	P15306 mus musculus
5	6	30.0	149	1 VLYS_BPPRD	P13559 bacterioph
6	6	30.0	215	1 ENGB_PSEAE	Q9ht81 pseudomonas
7	6	30.0	273	1 FPG_STRMU	P55045 streptococc
8	6	30.0	303	1 KGD_PSEPU	P42233 pseudomonas
9	6	30.0	327	1 COL7_ARATH	Q9sym2 arabidopsis
10	6	30.0	361	1 UXA1_BACHD	Q9kx28 bacillus ha
11	6	30.0	370	1 CTP1_PLAFK	P49587 plasmodium
12	6	30.0	508	1 HCM_PIG	O02734 sus scrofa
13	6	30.0	520	1 POLG_HCVH4	Q01404 hepatitis c
14	6	30.0	520	1 POLG_HCVHK	Q01403 hepatitis c
15	6	30.0	520	1 TX22_HUMAN	Q9Y458 homo sapien
16	6	30.0	601	1 MAON_SOLTU	P37225 solanum tub
17	6	30.0	737	1 POLG_HCVJ7	P27961 hepatitis c
18	6	30.0	1612	1 ATC4_YEAST	Q12675 saccharomyc
19	6	30.0	3010	1 POLG_HCVBK	P28663 h genome po
20	6	30.0	3010	1 POLG_HCVTW	P29846 h genome po
21	6	30.0	3033	1 POLG_HCVJ8	P26661 h genome po
22	5	25.0	73	1 YG62_YEAST	P53340 saccharomyc
23	5	25.0	92	1 FERN_AZOVI	P11054 azotobacter
24	5	25.0	105	1 RSNB_MOUSE	Q99p86 mus musculus
25	5	25.0	109	1 GLPE_YERPE	Q82j13 versinia pe
26	5	25.0	111	1 ESRI_SHEEP	P49885 ovis aries
27	5	25.0	111	1 TRKA_AZOCA	Q04856 azorhizobiu
28	5	25.0	113	1 Y182_AERPE	Q9Yfr7 aeropyrum p
29	5	25.0	114	1 RL42_PARAR	P41099 parthenium
30	5	25.0	120	1 R19E_THEAC	Q9h106 thermoplasm
31	5	25.0	140	1 VEG_HPV24	P50777 human papil
32	5	25.0	148	1 CT65_HUMAN	Q9h1h1 homo sapien
33	5	25.0	151	1 CT65_MOUSE	Q9cwd0 mus musculus

34	5	25.0	157	1	UBLG_HUMAN	P11441 homo sapien
35	5	25.0	157	1	UBLG_MOUSE	P21126 mus musculus
36	5	25.0	160	1	YFG8_YEAST	P43536 saccharomyc
37	5	25.0	167	1	RP07_FOWPV	Q05569 fowlpox vir
38	5	25.0	167	1	TFA_BPSFS	Q22005 bacterioph
39	5	25.0	175	1	XJ51_FUSNN	Q8rhq2 fuobacteri
40	5	25.0	176	1	YH92_AERPE	Q9Yb03 aeropyrum p
41	5	25.0	178	1	ATPD_BACST	P42008 bacillus st
42	5	25.0	180	1	ISPH_PSEFL	P21864 pseudomonas
43	5	25.0	183	1	PYRR_DEIRA	Q9rvb9 deinococcus
44	5	25.0	184	1	AROK_CHLTR	O84372 chlamydia t
45	5	25.0	185	1	NUSG_LACLA	Q9cdv7 lactococcus

ALIGNMENTS

RESULT 1

ID	POLG_HCV1	STANDARD;	PRT;	3011 AA.
AC	P26664;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].			
OS	Hepatitis C virus (isolate 1) (HCV).			
OC	Virusess ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.			
OX	NCBI_TaxID=11104;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91172826; PubMed=1848704;			
RA	Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Colt D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;			
RA	"Genetic organization and diversity of the hepatitis C virus.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).			
CC	-!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.			
CC	NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.			
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	-!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).			
CC	-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.			
CC	-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; M62321; AAA45676.1; -			
CC	PIR; A39166; GNMVCS.			
CC	PDB; 1A1V; 16-FEB-99.			
CC	PDB; 1HEI; 25-NOV-98.			
CC	MEROPS; S29.001; -			
CC	MEROPS; O39.001; -			
CC	InterPro; IPR001410; DEAD.			
CC	InterPro; IPR002522; HCV_capsid.			

InterPro; IPR002521; HCV core.
 InterPro; IPR002519; HCV env.
 InterPro; IPR002531; HCV NS1.
 InterPro; IPR002518; HCV NS2.
 InterPro; IPR004109; HCV NS3.
 InterPro; IPR000745; HCV NS4a.
 InterPro; IPR001490; HCV NS4b.
 InterPro; IPR002868; HCV NS5a.
 InterPro; IPR002166; HCV NS5a.
 InterPro; IPR001650; Helicase C.
 InterPro; IPR007095; RNA pol DS PS.
 InterPro; IPR007094; RNA pol PSvir.
 Pfam; PF01543; HCV capsid; 1.
 Pfam; PF01542; HCV core; 1.
 Pfam; PF01539; HCV env; 1.
 Pfam; PF01560; HCV NS1; 1.
 Pfam; PF01538; HCV NS2; 1.
 Pfam; PF02907; HCV NS3; 1.
 Pfam; PF01006; HCV NS4a; 1.
 Pfam; PF01001; HCV NS4b; 1.
 Pfam; PF01506; HCV NS5a; 1.
 Pfam; PF00271; Helicase C; 1.
 Pfam; PF00998; Viral RdRP; 1.
 Pfam; ProDom; PD186062; HCV NS1; 1.
 SMART; SM00487; DEXdc; 1.
 Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 3D-structure.
 INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 192 383 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 730 1006 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 1863 2013 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT CHAIN 2014 3011 POTENTIAL.
 FT TRANSMEM 347 369 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 ATP (POTENTIAL).
 FT NP_BIND 1230 1237 DECH BOX.
 FT SITE 1316 1319 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;
 Query Match 100.0%; Score 20; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 2.4e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GCPERLASCRPLTDFDQGWG 20

Db 451 GCPERLASCRPLTDFDQGWG 470
 RESULT 2
 POLG HCVJ2
 ID POLG HCVJ2 STANDARD; PRT; 513 AA.
 AC P27959;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 (GP68) (GP70) (NS1)] (Fragment).
 DE Hepatitis C virus (isolate HC-J2) (HCV).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11111;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: D10074; BAA00968.1; -;
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 DR Transmembrane; Nonstructural protein.
 KW INIT_MET 1 1
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 192 383 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
 FT CHAIN 384 513 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
 FT TRANSMEM 347 369 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT NON_TER 513 513

FT CARBOHYD 476 476 (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1224 1226 (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1232 1233
 FT TURN 1236 1238
 FT TURN 1239 1246
 FT HELIX 1247 1248
 FT STRAND 1251 1255
 FT STRAND 1258 1271
 FT HELIX 1272 1272
 FT TURN 1272 1280
 FT STRAND 1277 1280
 FT TURN 1281 1282
 FT STRAND 1283 1285
 FT STRAND 1291 1295
 FT HELIX 1296 1301
 FT TURN 1302 1303
 FT TURN 1312 1316
 FT TURN 1317 1319
 FT TURN 1323 1335
 FT TURN 1336 1340
 FT STRAND 1343 1347
 FT TURN 1352 1353
 FT TURN 1360 1361
 FT STRAND 1362 1366
 FT STRAND 1368 1368
 FT STRAND 1373 1375
 FT TURN 1376 1377
 FT STRAND 1378 1380
 FT HELIX 1382 1385
 FT STRAND 1389 1393
 FT HELIX 1397 1409
 FT TURN 1410 1411
 FT STRAND 1414 1417
 FT TURN 1419 1420
 FT STRAND 1432 1436
 FT TURN 1438 1439
 FT STRAND 1450 1453
 FT STRAND 1456 1463
 FT STRAND 1471 1478
 FT STRAND 1480 1480
 FT HELIX 1481 1488
 FT TURN 1489 1490
 FT STRAND 1497 1501
 FT STRAND 1507 1507
 FT STRAND 1511 1511
 FT STRAND 1514 1527
 FT HELIX 1532 1544
 FT STRAND 1550 1550
 FT HELIX 1555 1564
 FT STRAND 1570 1578
 FT TURN 1579 1580
 FT HELIX 1584 1597
 FT TURN 1598 1598
 FT HELIX 1606 1611
 FT TURN 1614 1618
 FT STRAND 1622 1623
 FT STRAND 1627 1627
 FT STRAND 1635 1636
 FT HELIX 1640 1652
 SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;

Query Match 50.0%; Score 10; DB 1; Length 3011;

Best Local Similarity 100.0%; Pred. No. 0.0089;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCR 10

Db 451 GCPERLASCR 460

RESULT 4

TRBM_MOUSE STANDARD; PRT; 577 AA.
 ID TRBM_MOUSE STANDARD; PRT; 577 AA.
 AC P15306;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thrombomodulin precursor (Fetomodulin) (TM).
 GN THBD.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89128454; PubMed=2536925;
 RA Dittman W.A., Majerus P.W.;
 RT "Sequence of a cDNA for mouse thrombomodulin and comparison of the
 predicted mouse and human amino acid sequences.";
 RL Nucleic Acids Res. 17:802-802(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89008498; PubMed=2844823;
 RA Dittman W.A., Kumada T., Sadler J.E., Majerus P.W.;
 RT "The structure and function of mouse thrombomodulin. Phorbol
 myristate acetate stimulates degradation and synthesis of
 thrombomodulin without affecting mRNA levels in hemangioma cells.";
 RL J. Biol. Chem. 263:15815-15822(1988).
 CC -1- FUNCTION: THROMBOMODULIN IS A SPECIFIC ENDOTHELIAL CELL RECEPTOR
 THAT FORMS A 1:1 STOICHIOMETRIC COMPLEX WITH THROMBIN. THIS
 COMPLEX IS RESPONSIBLE FOR THE CONVERSION OF PROTEIN C TO THE
 ACTIVATED PROTEIN C (PROTEIN CA). ONCE EVOLVED, PROTEIN CA
 SCISSIONS THE ACTIVATED COFACTORS OF THE COAGULATION MECHANISM,
 FACTOR VA AND FACTOR VIIIa, AND THEREBY REDUCES THE AMOUNT OF
 THROMBIN GENERATED.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: ENDOTHELIAL CELLS ARE UNIQUE IN SYNTHESIZING
 THROMBOMODULIN (BY SIMILARITY).
 CC -1- SIMILARITY: Contains 6 EGF-like domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X14432; CAA32597.1; -.
 DR PIR: S08488; A60501.
 DR HSSP: P07204; LEGT.
 DR MGD: MGI:98736; Thbd.
 DR GO: GO:0009790; P:embryonic development; IMP.
 DR GO: GO:0007565; P:pregnancy; IMP.
 DR InterPro: IPR000152; IPR001881; EGF_Ca.
 DR InterPro: IPR0001881; EGF_Ca.
 DR InterPro: IPR006209; EGF_like.
 DR InterPro: IPR001304; Lectin_C.
 DR InterPro: IPR001491; Thrombomoduln.
 DR Pfam: PF00008; EGF; 4.
 DR Pfam: PF00059; lectin_c; 1.
 DR PRINTS: PR00307; THROMBOMODULN.
 DR SMART: SM00034; CLECT; 1.
 DR SMART: SM00179; EGF_CA; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 2.
 DR PROSITE: PS00022; EGF_1; FALSE_NEG.
 DR PROSITE: PS01186; EGF_2; 3.
 DR PROSITE: PS01187; EGF_CA; 2.
 DR PROSITE: PS00041; C_TYPE_LECTIN_2; 1.
 DR Endothelial cell; Receptor; Blood coagulation; Repeat; Transmembrane;
 KW Glycoprotein; Signal; EGF-like domain.

FT SIGNAL 1 16 POTENTIAL.
 FT CHAIN 17 577 THROMBOMODULIN.
 FT DOMAIN 17 517 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 518 541 POTENTIAL.
 FT DOMAIN 542 577 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 240 280 EGF-LIKE 1.
 FT DOMAIN 283 323 EGF-LIKE 2.
 FT DOMAIN 324 362 EGF-LIKE 3. CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 363 404 EGF-LIKE 4.
 FT DOMAIN 403 439 EGF-LIKE 5.
 FT DOMAIN 440 480 EGF-LIKE 6. CALCIUM-BINDING (POTENTIAL).
 FT DISULFID 244 255 BY SIMILARITY.
 FT DISULFID 251 264 BY SIMILARITY.
 FT DISULFID 266 279 BY SIMILARITY.
 FT DISULFID 287 295 BY SIMILARITY.
 FT DISULFID 291 307 BY SIMILARITY.
 FT DISULFID 309 322 BY SIMILARITY.
 FT DISULFID 328 339 BY SIMILARITY.
 FT DISULFID 335 348 BY SIMILARITY.
 FT DISULFID 350 361 BY SIMILARITY.
 FT DISULFID 368 377 BY SIMILARITY.
 FT DISULFID 373 387 BY SIMILARITY.
 FT DISULFID 389 403 BY SIMILARITY.
 FT DISULFID 407 416 BY SIMILARITY.
 FT DISULFID 412 424 BY SIMILARITY.
 FT DISULFID 426 438 BY SIMILARITY.
 FT DISULFID 444 454 BY SIMILARITY.
 FT DISULFID 449 463 BY SIMILARITY.
 FT DISULFID 465 479 BY SIMILARITY.
 FT CARBOHYD 113 113 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 243 243 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 256 256 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 408 408 O-LINKED (GLYCOSAMINOGLYCAN) (BY
 FT CARBOHYD 494 494 SIMILARITY).
 SQ SEQUENCE 577 AA; 61867 MW; B20E50B0FE745014 CRC64;
 Query Match 35.0%; Score 7; DB 1; Length 577;
 Best Local Similarity 100.0%; Pred. No. 3.1;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 ASCRPLT 13
 Db 171 ASCRPLT 177
 RESULT 5
 ID VLYS BPPRD STANDARD; PRT; 149 AA.
 AC P13559;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 01-DEC-1992 (Rel. 24, Last annotation update)
 DE Lytic enzyme (Protein P15).
 GN XV.
 OS Bacteriophage PRD1.
 OC Viruses; dsDNA viruses, no RNA stage; Tectiviridae; Tectivirus.
 OX NCBI_TaxID=10658;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89210844; PubMed=2651121;
 RA Pakula T.M., Savilanti H., Bamford D.H.;
 RT "Comparison of the amino acid sequence of the lytic enzyme from
 RT broad-host-range bacteriophage PRD1 with sequences of other
 RT cell-wall-peptidoglycan lytic enzymes.";
 RL Eur. J. Biochem. 180:149-152(1989).
 RN [2]
 RP SEQUENCE OF 1-98 FROM N.A.
 RX MEDLINE=88112855; PubMed=3322943;
 RA Savilanti H., Bamford D.H.;
 RT "The complete nucleotide sequence of the left very early region of
 RT Escherichia coli bacteriophage PRD1 coding for the terminal protein
 RT and the DNA polymerase.";

RL Gene 57:121-130(1987).
 CC -!- FUNCTION: COMPLETE THE PRD1 LIFE CYCLE BY HOST CELL LYSIS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X14980; CAA33104.1; -;
 DR EMBL; M69077; AAA32457.1; -;
 DR EMBL; M22161; AAA32451.1; -;
 DR PIR; S03568; LYBPD1.
 KW Phage lysis protein.
 SQ SEQUENCE 149 AA; 17269 MW; 50ABB718032939B5 CRC64;
 Query Match 30.0%; Score 6; DB 1; Length 149;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 PERLAS 8
 Db 112 PERLAS 117
 RESULT 6
 ID ENGB_PSEAE STANDARD; PRT; 215 AA.
 AC Q9HT81;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable GTP-binding protein engb.
 GN ENGB OR PA5492.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Collier L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 CC -!- FUNCTION: NECESSARY FOR NORMAL CELL DIVISION AND FOR THE
 CC MAINTENANCE OF NORMAL SEPTATION (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE ENGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE004962; AAG08877.1; -;
 DR PIR; E82958; E82958.
 DR HAMAP; MF 00321; -;
 DR InterPro; IPR005289; GTP-binding_dom.
 DR TIGRFAMs; TIGR00650; MGA42; 1.
 KW Cell division; Septation; GTP-binding; Complete proteome.
 FT NP_BIND 38 45 GTP (POTENTIAL).
 FT NP_BIND 82 86 GTP (POTENTIAL).

```
FT NP_BIND 149 152 GTP (POTENTIAL).
SQ SEQUENCE 215 AA; 23786 MW; EC95DF81FA894A42 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 215;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 PLTDFD 16
Db 125 PLTDFD 130

RESULT 7
FPG_STRMU
ID FPG_STRMU STANDARD; PRT; 273 AA.
AC P55045;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Formamidopyrimidine-DNA glycosylase (EC 3.2.2.23) (FAPY-DNA
glycosylase).
GN MUTM OR FPG OR SMU.1614.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=11309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RA Takamatsu N.; Yamashita Y., Takehara T., Kuramitsu H.K.;
RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UAI59 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).
CC -!- FUNCTION: THIS ENZYME MAY PLAY A SIGNIFICANT ROLE IN PROCESSES
LEADING TO RECOVERY FROM MUTAGENESIS AND/OR CELL DEATH BY
ALKYLATING AGENTS.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of DNA containing ring-opened N(7)-
methylguanine residues, releasing 2,6-diamino-4-hydroxy-5-(N-
methyl)formamidopyrimidine.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE FPG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D26071; BAA05066.1; -.
DR EMBL; AE014992; AAN59255.1; -.
DR HSSP; O50606; 1EE8.
DR HAMAP; MF_00103; -.
DR InterPro; IPR000191; Fapy_DNA_glyco.
DR InterPro; IPR000214; Fapy_DNA_glyco_zn.
DR Pfam; PF01149; Fapy_DNA_glyco; 1.
DR ProDom; PD003680; Fapy_DNA_glyco; 1.
DR TIGRFAMs; TIGR00577; fpg; 1.
DR PROSITE; PS01242; FPG; 1.
KW DNA repair; Hydrolase; Glycosidase; Zinc; Zinc-finger;
Complete proteome.
ZN FING 247 270
FT CONFLICT 195 195 K -> T (IN REF. 1).
SQ SEQUENCE 273 AA; 30875 MW; 4A9ECE1C00D9C484 CRC64;
```

```
Query Match 30.0%; Score 6; DB 1; Length 273;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 PERLAS 8
Db 186 PERLAS 191

RESULT 8
KDGD_PSEPU
ID KDGD_PSEPU STANDARD; PRT; 303 AA.
AC P42233;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 5-dehydro-4-deoxyglucarate dehydratase (EC 4.2.1.41) (5-keto-4-deoxy-
glucarate dehydratase) (KDGDH).
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PP3;
RA Burlingame R.P., Lauer G.D., Platz J.G., Rudd E.A., Ally A.,
RA Ally D., Backman K.C.;
RT "Nucleotide sequence of genes for glucarate dehydratase and
5-keto-4-deoxyglucarate dehydratase from Pseudomonas putida pp3.";
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: 5-dehydro-4-deoxy-D-glucarate = 2,5-
dioxopentanoate + H(2)O + CO(2).
CC -!- SIMILARITY: TO THE B.SUBTILIS HOMOLOG.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M69160; AAA25869.1; -.
DR PIR; S27618; S27618.
DR InterPro; IPR002220; DHDPs.
DR Pfam; PF00701; DHDPs; 1.
DR PRINTS; PR00146; DHPICNTHASE.
DR ProDom; PD001859; DHDPs; 1.
KW Lyase.
SQ SEQUENCE 303 AA; 32808 MW; 9B9333EC14CF4D5E5 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 303;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 PLTDFD 16
Db 18 PLTDFD 23

RESULT 9
COL7_ARATH
ID COL7_ARATH STANDARD; PRT; 327 AA.
AC Q9SYW2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Zinc finger protein constans-like 7.
GN At1G78600 OR T30F21.7.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
```

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ev. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
 Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 Dunn P., Etnu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
 Militscher J., Miranda M., Nguyen M., Niernan W.C., Osborne B.I.,
 Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana.";
 RL Nature 408:816-820(2000).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- SIMILARITY: BELONGS TO THE CONSTANS FAMILY.
 CC -!- SIMILARITY: Contains 2 B box-type zinc fingers.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AC007260; AAD30576.1; -;
 DR PIR; F96814; F96814.
 DR InterPro; IPR000315; Znf Bbox.
 DR InterPro; IPR002926; Znf_constans.
 DR Pfam; PF00643; zF-B_box.2.
 DR ProDom; PD007661; Znf_constans.2.
 DR SMART; SM00336; BBOX.2.
 DR PROSITE; PS0119; ZF BBOX.2.
 KW Zinc-finger; Nuclear protein; Repeat; Multigene family.
 FT ZN_FING 5 47 B BOX-TYPE 1.
 FT ZN_FING 57 99 B BOX-TYPE 2.
 FT DOMAIN 149 155 POLY-GLN.
 SQ SEQUENCE 327 AA; 35932 MW; C8683E44D0504D6 CRC64;
 Query Match 30.0%; Score 6; DB 1; Length 327;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 12 LTDFDQ 17
 DB 195 LTDFDQ 200
 RESULT 10
 ID UXAL BACHD STANDARD; PRT; 361 AA.
 AC Q9KDX8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Mannonate dehydratase 1 (EC 4.2.1.8) (D-mannonate hydrolase 1).
 GN UXAL1 OR BH1063
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=86665;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";
 RL Nucleic Acids Res. 28:4317-4331(2000).
 CC -!- CATALYTIC ACTIVITY: D-mannonate = 2-dehydro-3-deoxy-D-gluconate +
 CC H₂O.
 CC -!- PATHWAY: Glucuronate pathway.
 CC -!- SIMILARITY: Belongs to the mannose dehydratase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AP001510; BAB04782.1; -;
 DR PIR; G83782; G83782.
 DR HAMAP; MF 00106; -; 1.
 DR InterPro; IPR004628; Man_dehyd.
 DR Pfam; PF03786; Uxua.1.
 DR TIGRFAMs; TIGR00695; uxua.1.
 KW Lyase; Complete proteome.
 SQ SEQUENCE 361 AA; 41633 MW; 4C45A43278E1E732 CRC64;
 Query Match 30.0%; Score 6; DB 1; Length 361;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 3 PERLAS 8
 DB 160 PERLAS 165
 RESULT 11
 ID CPTPLAFLK STANDARD; PRT; 370 AA.
 AC P49587;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cholinephosphate cytidylyltransferase (EC 2.7.7.15) (Phosphorylcholine
 DE transferase) (CTP:phosphocholine cytidylyltransferase) (CT) (CCT).
 GN CTP.
 OS Plasmodium falciparum (isolate K1 / Thailand).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5839;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96061933; PubMed=7588775;
 RA Yeo H.-J., Sri Widada J., Mercereau-Puijalon O., Vial H.J.;
 RT "Molecular cloning of CTP:phosphocholine cytidylyltransferase from
 RT Plasmodium falciparum.";
 RL Eur. J. Biochem. 233:62-72(1995).
 CC -!- FUNCTION: CONTROLS PHOSPHATIDYLCHOLINE SYNTHESIS.
 CC -!- CATALYTIC ACTIVITY: CTP + choline phosphate = diphosphate + CDP-
 CC choline.
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE ASEXUAL INTRACERYTHROCYTIC
 CC STAGES.
 CC -!- SIMILARITY: Belongs to the cytidylyltransferase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

or send an email to license@isb-sib.ch).

EMBL; X84041; CRA58860.1; -
 PIR; S68187; S68187.
 HSP; P27623; 1C0Z.
 InterPro; IPR004821; Cyt_tran_rel.
 InterPro; IPR004820; Cytidyltransf.
 Pfam; PF01467; CTP_transf.2; 1.
 TIGRFAMs; TIGR00125; Cyt_tran_rel; 1.
 Transferase; Nucleosidyltransferase; Phospholipid biosynthesis.
 FT DOMAIN 89 268 CATALYTIC (POTENTIAL).
 FT DOMAIN 57 64 POLY-ASP.
 FT DOMAIN 193 203 POLY-LYS.
 SQ SEQUENCE 370 AA; 42630 MW; B2E457D09B0B6FB6 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 370;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 TDFDQG 18.
 DB 302 TDFDQG 307
 |||||

RESULT 12
 HNCM_PIG STANDARD; PRT; 508 AA.
 AC 002734;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hydroxymethylglutaryl-CoA synthase, mitochondrial precursor
 DE (EC 2.3.3.10) (HMG-CoA synthase) (3-hydroxy-3-methylglutaryl coenzyme
 A synthase).
 GN HMGCS2.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=97293201; PubMed=9164842;
 RA Adams S.H., Alho C.S., Asins G., Hegardt F.G., Marrero P.F.;
 RT "Gene expression of mitochondrial 3-hydroxy-3-methylglutaryl-CoA
 synthase in a poorly ketogenic mammal: effect of starvation during
 the neonatal period of the piglet.";
 RL Biochem. J. 324:65-73(1997).
 CC -!- FUNCTION: THIS ENZYME CONDENSES ACETYL-CoA WITH ACETOACETYL-CoA
 TO FORM HMG-CoA, WHICH IS THE SUBSTRATE FOR HMG-CoA REDUCTASE.
 CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
 hydroxy-3-methylglutaryl-CoA + CoA.
 CC -!- PATHWAY: RESPONSIBLE, TOGETHER WITH HMG-CoA LYASE FOR KETONE
 BODY BIOSYNTHESIS.
 CC -!- SUBCELLULAR LOCATION: Mitochondrial.
 CC -!- SIMILARITY: Belongs to the HMG-CoA synthase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; U90884; AAC48727.1; -
 InterPro; IPR000590; HMG CoA synt.
 Pfam; PF01154; HMG CoA synt.; 1.
 PROSITE; PS01226; HMG CoA SYNTHASE; 1.
 Transferase; Cholesterol Biosynthesis; Mitochondrion; Transit peptide;
 Multigene family.
 FT TRANSIT 1 37
 FT CHAIN 38 508 MITOCHONDRION (PROBABLE).
 FT CARBOHYD 38 508 HYDROXYMETHYLGLUTARYL-CoA SYNTHASE.

FT ACT SITE 166 166 POTENTIAL.
 SQ SEQUENCE 508 AA; 56933 MW; 5479DE6F70B3C0F6 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 508;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PERLAS 8
 DB 446 PERLAS 451
 |||||

RESULT 13
 POLG_HCVH4 STANDARD; PRT; 520 AA.
 AC 001404;
 DT 01-JUN-1993 (Rel. 26, Created)
 DT 01-JUN-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1)] (Fragment).
 OS Hepatitis C virus (isolate HCV-476) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31643;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93019030; PubMed=1383400;
 RA Abe K., Inchauspe G., Fujisawa K.;
 RT "Genomic characterization and mutation rate of hepatitis C virus
 isolated from a patient who contracted hepatitis during an epidemic
 of non-A, non-B hepatitis in Japan.";
 RL J. Gen. Virol. 73:2725-2729(1992).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; D10688; BAA01530.1; -
 InterPro; IPR002522; HCV capsid.
 InterPro; IPR002521; HCV core.
 InterPro; IPR002519; HCV env.
 InterPro; IPR002531; HCV_NSI.
 Pfam; PF01543; HCV capsid; 1.
 Pfam; PF01542; HCV core; 1.
 Pfam; PF01539; HCV env; 1.
 Pfam; PF01560; HCV_NSI; 1.
 ProDom; PD186062; HCV_NSI; 1.
 Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 Transmembrane; Nonstructural protein.
 KW Transmembrane; Nonstructural protein.
 FT INIT_MET 1 1
 FT CHAIN 1 115 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 116 191 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 192 383 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 384 520 MATRIX PROTEIN (POTENTIAL).
 FT TRANSMEM 347 369 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CARBOHYD 196 196 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CARBOHYD 209 209 BY SIMILARITY.
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON TER 520 520
 SQ SEQUENCE 520 AA; 56499 MW; AA135246CF20D525 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 520;

Best Local Similarity 100.0%; Pred.No. 32; Indels 0; Gaps 0;
 Matches 6; Conservative 0; Mismatches 0;

Qy 15 FDOGWG 20

Db 466 FDOGWG 471

RESULT 14

POLG_HCVHK STANDARD; PRT; 520 AA.
 ID POLG_HCVHK STANDARD; PRT; 520 AA.
 AC Q01403;
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1)) (Fragment).
 OS Hepatitis C virus (isolate HCV-KP) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31644;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93019030; PubMed=1383400;
 RA Abe K., Inchauspe G., Fujisawa K.;
 RT "Genomic characterization and mutation rate of hepatitis C virus
 isolated from a patient who contracted hepatitis during an epidemic
 of non-A, non-B hepatitis in Japan.";
 RT J. Gen. Virol. 73:2725-2728(1992).
 RL J. Gen. Virol. 73:2725-2728(1992).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND RNA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; D10687; BAA01529.1; -
 PIR; JQ1925; JQ1925.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Prodom; PDI86062; HCV NS1; 1.
 DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 KX
 FT INIT_MET 1
 FT 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 >520
 FT TRANSMEM 347 369
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305

EMBL; D10687; BAA01529.1; -
 PIR; JQ1925; JQ1925.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Prodom; PDI86062; HCV NS1; 1.
 DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 KX
 FT INIT_MET 1
 FT 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 >520
 FT TRANSMEM 347 369
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305

FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON TER 520 520
 SQ SEQUENCE 520 AA; 56476 MW; 1D2BD0A6FF27349B CRC64;

Query Match 30.0%; Score 6; DB 1; Length 520;

Best Local Similarity 100.0%; Pred.No. 32; Indels 0; Gaps 0;
 Matches 6; Conservative 0; Mismatches 0;

Qy 15 FDOGWG 20

Db 466 FDOGWG 471

RESULT 15

TX22_HUMAN STANDARD; PRT; 520 AA.
 ID TX22_HUMAN STANDARD; PRT; 520 AA.
 AC Q9Y458; Q96LC0; Q9HBF1;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE T-box transcription factor TBX22 (T-box protein 22).
 GN TBX22 OR TBOX22.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANTS CPX CVS-118 AND MET-260.
 RX MEDLINE=21470334; PubMed=11559848;
 RA Braybrook C., Doudney K., Marciano A.C., Arnason A., Bjornsson A.,
 RA Patton M.A., Goodfellow P.J., Moore G.E., Stanier P.;
 RT "The T-box transcription factor gene TBX22 is mutated in X-linked
 cleft palate and ankyloglossia.";
 RT Nat. Genet. 29:179-183(2001).
 RL Nat. Genet. 29:179-183(2001).
 RN [2]
 RP SEQUENCE OF 121-520 FROM N.A.
 RX MEDLINE=20480696; PubMed=11024289;
 RA Laugier-Anfossi F., Villard L.;
 RT "Molecular characterization of a new human T-box gene (TBX22) located
 in Xq21.1 encoding a protein containing a truncated T-domain.";
 RL Gene 255:289-296(2000).
 RN [3]
 RP SEQUENCE OF 96-287 FROM N.A.
 RA Pearce A.;
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Probable transcriptional regulator involved in
 developmental processes. This is major determinant crucial to
 palatogenesis.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- TISSUE SPECIFICITY: Seems to be expressed at a low level.
 CC -1- DISEASE: Defects in TBX22 are the cause of X-linked cleft palate
 with ankyloglossia (CPX) [MIM:303400].
 CC -1- SIMILARITY: Contains 1 T-box domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; AY035371; AAK63189.1; -
 EMBL; AF251684; AAG23749.1; -
 EMBL; AL031000; CAB38835.1; -
 HSSP; P24781; 1XER.
 TRANSFAC; T04949; -
 DR GeneW; HGNC:11600; TBX22.
 DR MIN; 300307; -
 DR MIN; 303400; -

DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0007275; P:development; TAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; TAS.
DR InterPro; IPR001699; TF_T-box.
DR Pfam; PF00907; T-box; 1.
DR PRINTS; PR00937; TBOX.
DR SMART; SM00425; TBOX; 1.
DR PROSITE; PS01283; TBOX_1; 1.
DR PROSITE; PS01264; TBOX_2; 1.
DR PROSITE; PS0252; TBOX_3; 1.
KW Transcription regulation; DNA-binding; Nuclear protein;
Disease mutation.
FT DNA_BIND 96 283 T-BOX.
FT VARIANT 118 118 G -> C (in CPX).
FT VARIANT 260 260 /FTid=VAR_015383.
FT VARIANT 260 260 T -> M (in CPX).
FT /FTid=VAR_015384.
SQ SEQUENCE 520 AA; 57910 MW; 1502E19245CC2C93 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 520;
Best Local Similarity 100.0%; Pred.No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 PERLAS 8
| | | | |
Db 391 PERLAS 396

Search completed: November 21, 2003, 20:59:54
Job time : 5.85 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:45:57 ; Search time 24.4 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-76

Perfect score: 20

Sequence: 1 GCPERLASCRPLTDFDQGWG 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	116	12 Q81443	Q81443 hepatitis c
2	20	100.0	123	12 Q81442	Q81442 hepatitis c
3	20	100.0	130	12 Q912L4	Q912L4 hepatitis c
4	20	100.0	130	12 Q912L7	Q912L7 hepatitis c
5	20	100.0	130	12 Q912L8	Q912L8 hepatitis c
6	20	100.0	130	12 Q912L6	Q912L6 hepatitis c
7	20	100.0	130	12 Q912L3	Q912L3 hepatitis c
8	20	100.0	130	12 Q912L5	Q912L5 hepatitis c
9	20	100.0	130	12 Q912L9	Q912L9 hepatitis c
10	20	100.0	135	12 Q9YJA3	Q9YJA3 hepatitis c
11	20	100.0	135	12 Q919S4	Q919S4 hepatitis c
12	20	100.0	135	12 Q9YK84	Q9YK84 hepatitis c
13	20	100.0	138	12 Q81453	Q81453 hepatitis c
14	20	100.0	169	12 Q911U4	Q911U4 hepatitis c
15	20	100.0	169	12 Q917R4	Q917R4 hepatitis c
16	20	100.0	169	12 Q917R5	Q917R5 hepatitis c

17	20	100.0	169	12 Q911U3	Q911U3 hepatitis c
18	20	100.0	169	12 Q917R8	Q917R8 hepatitis c
19	20	100.0	169	12 Q917R9	Q917R9 hepatitis c
20	20	100.0	179	12 Q8QR84	Q8QR84 hepatitis c
21	20	100.0	190	12 Q8JPE2	Q8JPE2 hepatitis c
22	20	100.0	191	12 Q8JPD1	Q8JPD1 hepatitis c
23	20	100.0	191	12 Q8JPC8	Q8JPC8 hepatitis c
24	20	100.0	191	12 Q8JPD8	Q8JPD8 hepatitis c
25	20	100.0	191	12 Q8JPF0	Q8JPF0 hepatitis c
26	20	100.0	191	12 Q8JPE3	Q8JPE3 hepatitis c
27	20	100.0	191	12 Q8JPE0	Q8JPE0 hepatitis c
28	20	100.0	191	12 Q8JPD0	Q8JPD0 hepatitis c
29	20	100.0	191	12 Q8JPD6	Q8JPD6 hepatitis c
30	20	100.0	191	12 Q8JPE1	Q8JPE1 hepatitis c
31	20	100.0	191	12 Q8JPD2	Q8JPD2 hepatitis c
32	20	100.0	191	12 Q8JPD4	Q8JPD4 hepatitis c
33	20	100.0	191	12 Q8JPD5	Q8JPD5 hepatitis c
34	20	100.0	191	12 Q8JPF2	Q8JPF2 hepatitis c
35	20	100.0	191	12 Q8JPE4	Q8JPE4 hepatitis c
36	20	100.0	191	12 Q8JPD9	Q8JPD9 hepatitis c
37	20	100.0	191	12 Q8JPC7	Q8JPC7 hepatitis c
38	20	100.0	191	12 Q8JPE5	Q8JPE5 hepatitis c
39	20	100.0	191	12 Q8BKK0	Q8BKK0 hepatitis c
40	20	100.0	191	12 Q8BBJ9	Q8BBJ9 hepatitis c
41	20	100.0	191	12 Q8BBJ6	Q8BBJ6 hepatitis c
42	20	100.0	191	12 Q8BBJ5	Q8BBJ5 hepatitis c
43	20	100.0	191	12 Q8BBJ3	Q8BBJ3 hepatitis c
44	20	100.0	325	12 Q9WG32	Q9WG32 hepatitis c
45	20	100.0	326	12 Q98UV3	Q98UV3 hepatitis c

ALIGNMENTS

RESULT 1

Q81443 ID Q81443 PRELIMINARY; PRT; 116 AA.
AC Q81443;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Unknown protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CG2-5;
RA Hall W.W., Suzuki H.;
RT "Genetic heterogeneity of Hepatitis C viruses in a population of HIV infected intravenous drug abusers (IVDAs).";
RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L19372; AAA45593.1; -.
DR InterPro; IPR002531; HCV.NS1.
DR Pfam; PF01560; HCV.NS1; 1.
DR ProDom; PD186062; HCV.NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON TER 1
FT NON TER 116
SQ SEQUENCE 116 AA; 12761 MW; 57019C06E738F739 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 116;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GCPERLASCRPLTDFDQGWG 20

Db 76 GCPERLASCRPLTDFDQGWG 95

RESULT 2

Qy	1	GCPERLASCRPLTDFDQGWG	20	
Db	79	GCPERLASCRPLTDFDQGWG	98	
RESULT 4				
Qy	Q9IZL7	PRELIMINARY;	PRT;	130 AA.
AC	Q9IZL7;			
DT	01-OCT-2000 (TrEMBLrel. 15, Created)			
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)			
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)			
DE	Genome polypeptide (Fragment).			
OS	Hepatitis C virus.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC	Hepacivirus.			
NCBI	TaxID=111103;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=X304A-G;			
RX	MEDLINE=20173694; PubMed=10708420;			
RA	Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,			
RA	Wang Y.M., Thomas D.L.;			
RT	"Hypervariable region 1 sequence stability during hepatitis C virus			
RT	replication in chimpanzees.";			
RL	J. Virol. 74:3058-3066(2000).			
RL	EMBL; AF230455; AAF68723.1; -.			
DR	InterPro; IPR002531; HCV NS1.			
DR	Pfam; PF01560; HCV NS1; 1.			
DR	ProDom; PD186062; HCV NS1; 1.			
KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein			
KW	Polypeptide; Transmembrane.			
FT	NON_TER 1			
FT	NON_TER 130 130			
SQ	SEQUENCE 130 AA; 13811 MW; 5245C05292F3B659 CRC64;			
Query Match 100.08; Score 20; DB 12; Length 130;				
Best Local Similarity 100.08; Pred. No. 1.3e-13;				
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps				
Qy	1	GCPERLASCRPLTDFDQGWG	20	
Db	79	GCPERLASCRPLTDFDQGWG	98	
RESULT 5				
Qy	Q9IZL8	PRELIMINARY;	PRT;	130 AA.
ID	Q9IZL8 o			
AC	Q9IZL8;			
DT	01-OCT-2000 (TrEMBLrel. 15, Created)			
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)			
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)			
DE	Genome polypeptide (Fragment).			
OS	Hepatitis C virus.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC	Hepacivirus.			
NCBI	TaxID=111103;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=X304A-B;			
RX	MEDLINE=20173694; PubMed=10708420;			
RA	Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,			
RA	Wang Y.M., Thomas D.L.;			
RT	"Hypervariable region 1 sequence stability during hepatitis C virus			
RT	replication in chimpanzees.";			
RL	J. Virol. 74:3058-3066(2000).			
RL	EMBL; AF230454; AAF68722.1; -.			
DR	InterPro; IPR002531; HCV NS1.			
DR	Pfam; PF01560; HCV NS1; 1.			
DR	ProDom; PD186062; HCV NS1; 1.			
KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein			
KW	Polypeptide; Transmembrane.			

```
FT  NON TER      1      1
FT  NON TER     130     130
SQ  SEQUENCE    130 AA; 13801 MW;  ABAD711822F3A2A3 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 130;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 GCPERLASCRPLTDFDQGWG 20
    |||||
DB  79 GCPERLASCRPLTDFDQGWG 98

RESULT 6
Q9IZL6
ID  Q9IZL6      PRELIMINARY;      PRT;      130 AA.
AC  Q9IZL6;
DT  01-OCT-2000 (TrEMBLrel. 15, Created)
DT  01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE  Genome polyprotein (Fragment).
OS  Hepatitis C virus.
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC  Hepacivirus.
OX  NCBI_TaxID=111103;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=X304C-I;
RX  MEDLINE=20173694; PubMed=10708420;
RA  Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RA  Wang Y.M., Thomas D.L.;
RT  "Hypervariable region 1 sequence stability during hepatitis C virus
RT  replication in chimpanzees.";
RL  J. Virol. 74:3058-3066(2000).
DR  EMBL; AF230456; AAF68724.1; -.
DR  InterPro; IPR002531; HCV NS1.
DR  Pfam; PF01560; HCV NS1; 1.
DR  ProDom; PD186062; HCV NS1; 1.
KW  Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW  Polyprotein; Transmembrane.
FT  NON TER      1
FT  NON TER     130     130
SQ  SEQUENCE    130 AA; 13899 MW;  0B6A41E9E331F75F CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 130;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 GCPERLASCRPLTDFDQGWG 20
    |||||
DB  79 GCPERLASCRPLTDFDQGWG 98

RESULT 7
Q9IZL3
ID  Q9IZL3      PRELIMINARY;      PRT;      130 AA.
AC  Q9IZL3;
DT  01-OCT-2000 (TrEMBLrel. 15, Created)
DT  01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE  Genome polyprotein (Fragment).
OS  Hepatitis C virus.
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC  Hepacivirus.
OX  NCBI_TaxID=111103;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=X304C-U;
RX  MEDLINE=20173694; PubMed=10708420;
RA  Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RA  Wang Y.M., Thomas D.L.;
RT  "Hypervariable region 1 sequence stability during hepatitis C virus
RT  replication in chimpanzees.";
```

```
J. Virol. 74:3058-3066(2000).
DR  EMBL; AF230459; AAF68727.1; -.
DR  InterPro; IPR002531; HCV NS1.
DR  Pfam; PF01560; HCV NS1; 1.
DR  ProDom; PD186062; HCV NS1; 1.
KW  Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW  Polyprotein; Transmembrane.
FT  NON TER      1
FT  NON TER     130     130
SQ  SEQUENCE    130 AA; 13827 MW;  0B692289E331F55D CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 130;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 GCPERLASCRPLTDFDQGWG 20
    |||||
DB  79 GCPERLASCRPLTDFDQGWG 98

RESULT 8
Q9IZL5
ID  Q9IZL5      PRELIMINARY;      PRT;      130 AA.
AC  Q9IZL5;
DT  01-OCT-2000 (TrEMBLrel. 15, Created)
DT  01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE  Genome polyprotein (Fragment).
OS  Hepatitis C virus.
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC  Hepacivirus.
OX  NCBI_TaxID=111103;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=X304C-K;
RX  MEDLINE=20173694; PubMed=10708420;
RA  Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RA  Wang Y.M., Thomas D.L.;
RT  "Hypervariable region 1 sequence stability during hepatitis C virus
RT  replication in chimpanzees.";
RL  J. Virol. 74:3058-3066(2000).
DR  EMBL; AF230457; AAF68725.1; -.
DR  InterPro; IPR002531; HCV NS1.
DR  Pfam; PF01560; HCV NS1; 1.
DR  ProDom; PD186062; HCV NS1; 1.
KW  Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW  Polyprotein; Transmembrane.
FT  NON TER      1
FT  NON TER     130     130
SQ  SEQUENCE    130 AA; 13827 MW;  0B692289E331F55D CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 130;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 GCPERLASCRPLTDFDQGWG 20
    |||||
DB  79 GCPERLASCRPLTDFDQGWG 98

RESULT 9
Q9IZL9
ID  Q9IZL9      PRELIMINARY;      PRT;      130 AA.
AC  Q9IZL9;
DT  01-OCT-2000 (TrEMBLrel. 15, Created)
DT  01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE  Genome polyprotein (Fragment).
OS  Hepatitis C virus.
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC  Hepacivirus.
OX  NCBI_TaxID=111103;
RN  [1]
```

```

RP SEQUENCE FROM N.A.
RC STRAIN=X304A-A;
RX MEDLINE=20173694; PubMed=10708420;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
  Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
  replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL; AF230453; AAF68721.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
  Polyprotein; Transmembrane.
KW Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
FT NON_TER 130 130
FT NON_TER 130 130
SQ SEQUENCE 130 AA; 13783 MW; 69287689B661F558 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
Db 79 GCPERLASCRPLTDFDQGWG 98

RESULT 10
Q9YJA3 PRELIMINARY; PRT; 135 AA.
AC Q9YJA3;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411414; PubMed=9738054;
RA Wang Y.M., Ray S.C., Laeyendecker O., Ticehurst J.R., Thomas D.L.;
RT "Assessment of hepatitis C virus sequence complexity by
  electrophoretic mobilities of both single- and double-stranded DNAs.";
RL J. Clin. Microbiol. 36:2982-2989(1998).
DR EMBL; AF073023; AAC61323.1; -.
DR EMBL; AF073021; AAC61321.1; -.
DR InterPro; IPR002531; HCV env.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
  Polyprotein; Transmembrane.
KW Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
FT NON_TER 1 1
FT CHAIN 1 >22 E1.
FT CHAIN 23 >135 E2.
FT NON_TER 135 135
SQ SEQUENCE 135 AA; 14549 MW; 67FF22166B0B846 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 135;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
Db 90 GCPERLASCRPLTDFDQGWG 109

RESULT 11
Q91954 PRELIMINARY; PRT; 135 AA.
AC Q91954;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411414; PubMed=9738054;
RA Wang Y.M., Ray S.C., Laeyendecker O., Ticehurst J.R., Thomas D.L.;
RT "Assessment of hepatitis C virus sequence complexity by
  electrophoretic mobilities of both single- and double-stranded DNAs.";
RL J. Clin. Microbiol. 36:2982-2989(1998).
DR EMBL; AF073023; AAC61323.1; -.
DR EMBL; AF073021; AAC61321.1; -.
DR InterPro; IPR002531; HCV env.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
  Polyprotein; Transmembrane.
KW Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
FT NON_TER 1 1
FT CHAIN 1 >22 E1.
FT CHAIN 23 >135 E2.
FT NON_TER 135 135
SQ SEQUENCE 135 AA; 14549 MW; 7A22FD36BEC6BF76 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 135;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
Db 90 GCPERLASCRPLTDFDQGWG 109

```

```

AC Q91954;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411414; PubMed=9738054;
RA Wang Y.M., Ray S.C., Laeyendecker O., Ticehurst J.R., Thomas D.L.;
RT "Assessment of hepatitis C virus sequence complexity by
  electrophoretic mobilities of both single- and double-stranded DNAs.";
RL J. Clin. Microbiol. 36:2982-2989(1998).
DR EMBL; AF073020; AAC61320.1; -.
DR EMBL; AF073024; AAC61324.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
  Polyprotein; Transmembrane.
KW Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
FT NON_TER 1 1
FT NON_TER 135 135
SQ SEQUENCE 135 AA; 14551 MW; 7A22E20BA1FBA846 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 135;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
Db 90 GCPERLASCRPLTDFDQGWG 109

RESULT 12
Q9YK84 PRELIMINARY; PRT; 135 AA.
AC Q9YK84;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411414; PubMed=9738054;
RA Wang Y.M., Ray S.C., Laeyendecker O., Ticehurst J.R., Thomas D.L.;
RT "Assessment of hepatitis C virus sequence complexity by
  electrophoretic mobilities of both single- and double-stranded DNAs.";
RL J. Clin. Microbiol. 36:2982-2989(1998).
DR EMBL; AF073022; AAC61322.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
  Polyprotein; Transmembrane.
KW Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
FT NON_TER 1 1
FT CHAIN 1 >22 E1.
FT CHAIN 23 >135 E2.
FT NON_TER 135 135
SQ SEQUENCE 135 AA; 14549 MW; 7A22FD36BEC6BF76 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 135;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCPERLASCRPLTDFDQGWG 20
Db 90 GCPERLASCRPLTDFDQGWG 109

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:22 ; Search time 32.1 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-77

Perfect score: 20

Sequence: 1 TDFDQGWGPIYSYANGSGPDQ 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A Geneseq 19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	16 AAR84506	Hepatitis C virus
2	20	100.0	20	17 AAR91000	HCV E2 peptide E2-
3	20	100.0	20	23 AAO18706	Hepatitis C virus
4	20	100.0	37	12 AAR13359	p447 HCV antigen (
5	20	100.0	37	14 AAR33878	Polypeptide p447 c
6	20	100.0	76	21 AAB18528	Protein encoded by
7	20	100.0	254	22 AAB68043	Amino acid sequenc
8	20	100.0	278	14 AAR33997	Th E2/NS1 protein.
9	20	100.0	363	22 AAB68042	Amino acid sequenc

10	20	100.0	402	14 AAR34438	Sequence of glycop
11	20	100.0	454	10 AAP90183	Sequence of hepati
12	20	100.0	454	10 AAP92049	Sequence encoded b
13	20	100.0	454	21 AAB18526	Protein encoded by
14	20	100.0	480	14 AAR33992	HCV-1 E2/NS1 prote
15	20	100.0	531	22 AAE02622	Chimeric HCV E2661
16	20	100.0	637	24 ABP57410	Hepatitis C virus
17	20	100.0	663	17 AAR92935	HCV1 E2 + NS2 poly
18	20	100.0	663	20 AAW67615	Hepatitis C virus
19	20	100.0	2435	13 AAR251135	HCV polypeptide 1.
20	20	100.0	2436	10 AAP92050	Sequence encoded i
21	20	100.0	2436	10 AAP90288	Peptide encoded by
22	20	100.0	2436	13 AAR28582	HCV amino acid seq
23	20	100.0	2772	11 AAR08123	Hepatitis C virus
24	20	100.0	2772	21 AAB18540	Protein encoded by
25	20	100.0	2816	14 AAR34009	HCV-1 polyprotein.
26	20	100.0	2894	13 AAR24440	Composite HCV HC-J
27	20	100.0	2894	16 AAR70230	Composite hepatiti
28	20	100.0	2955	11 AAR08124	Hepatitis C virus
29	20	100.0	2955	20 AAY14975	Amino acid sequenc
30	20	100.0	2955	21 AAB18541	Polyprotein encode
31	20	100.0	3011	13 AAR21519	Compiled HCV seque
32	20	100.0	3011	14 AAR31621	Hepatitis C virus
33	20	100.0	3011	17 AAR90931	Hepatitis C virus
34	20	100.0	3011	18 AAW34480	HCV polyprotein.
35	20	100.0	3011	19 AAW40038	Hepatitis C virus
36	20	100.0	3011	23 AAE22049	Hepatitis C virus
37	20	100.0	3011	23 AAE22052	Hepatitis C virus
38	20	100.0	3011	23 AAU84597	HCV polyprotein la
39	19	95.0	30	23 AAU84629	HCV HepCia segment
40	19	95.0	402	14 AAR34440	Sequence of glycop
41	19	95.0	409	14 AAR33996	H90 E2/NS1 protein
42	19	95.0	2010	23 AAU84801	HCV HepC cassette
43	19	95.0	5985	23 AAU84799	HCV HepCia scrambl
44	18	90.0	192	19 AAW67009	HCV nucleocapsid c
45	16	80.0	27	14 AAR34060	HCV E2 domain pept

ALIGNMENTS

RESULT 1
AAR84506
ID AAR84506 standard; peptide; 20 AA.
AC AAR84506;
XX
XX
DT 06-JAN-1997 (first entry)
XX
XX
DE Hepatitis C virus peptide NS1-3 (residues 463-482).
XX
XX
KW Hepatitis C virus; HCV; immunogen; non-structural region; NS1;
XX immunodominant; T cell epitope; vaccine.
OS Hepatitis C virus.
XX
XX
FN WO9512677-A2.
XX
PD 11-MAY-1995.
XX
PF 28-OCT-1994; 94WO-EP03555.
XX
PR 04-NOV-1993; 93EP-0402718.
XX
XX (INNO-) INNOGENETICS NV.
XX
PI Deleys R, Leroux-Roels G, Maertens G;
XX
DR WPI; 1995-193822/25.
XX
PT Hepatitis C Virus immunogenic polypeptide contg. a T-cell.
PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in
PT production of vaccines, therapeutic agents, etc.

XX PS Example 4; Page 51; 105pp; English.

XX CC A series of overlapping peptides (including the present sequence) was

CC synthesised based on sequences in the core, E1 and E2/NS1 regions of

CC hepatitis C virus. The peptides were used as antigens in lympho-

CC proliferative assays to identify the main T-cell epitopes.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 16; Length 20;

Best Local Similarity 100.0%; Pred. No. 3.3e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20

Db 1 TDFDQGWGPISYANGSGPDQ 20

RESULT 2

AAR91000

ID AAR91000 standard; peptide; 20 AA.

XX AC AAR91000;

XX DT 25-SEP-1996 (first entry)

XX DE HCV E2 peptide E2-3B for competition studies.

XX KW HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;

XX KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.

XX OS Synthetic.

XX PN WO9604385-A2.

XX PD 15-FEB-1996.

XX PF 31-JUL-1995; 95WO-EP03031.

XX PR 29-JUL-1994; 94EP-0870132.

XX PA (INNO-) INNOGENETICS NV.

XX PI Bosman F, Buyse M, De Martynoff G, Maertens G;

XX DR WPI; 1996-129401/13.

XX PT Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope

PT proteins - in presence of disulphide bond cleavage agent, to

PT produce proteins suitable for direct use in vaccines or diagnostic

PT assays of HCV

XX PS Example 7; Page 67; 146pp; English.

XX CC AAR90965-R90980 and AAR90995-R91015 represent synthetic hepatitis C

CC virus (HCV) E1 and E2 peptides used in competition studies. This

CC sequence represents a synthetic E2 peptide, and corresponds to residues

CC 463-482 of the E2 protein sequence. These sequences are useful for in

CC vitro monitoring of HCV disease, or prognosis of the response to

CC interferon treatment of patients suffering from HCV infection. These

CC sequences compete with the proteins produced by AAT12704-T12709 and

CC AAT12961-T12974, which are included in vectors for the production of

CC recombinant E1, E2, and E1/E2. The recombinant proteins can then be

CC isolated and purified by carrying out a disulphide bond cleavage, or a

CC reduction step with a disulphide bond cleavage agent, after lysis of

CC recombinant host cells. The constructs containing the purified HCV

CC envelope proteins can be used for vaccinating humans against HCV, for in

CC vitro detection of HCV antibodies in a sample, and in a serotyping assay

CC for detecting one or more serological types of HCV present in a

CC biological sample. The constructs can also be immobilised on a solid

CC substrate and incorporated into a reversed phase hybridisation assay for

CC determining the presence or the genotype of HCV. The new purification

CC method preserves the conformation of the recombinantly expressed E1, E2

CC and E1/E2, and eliminates contaminating proteins. Antigens isolated

CC using this method are more reactive with human sera than those isolated

CC by known techniques.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 17; Length 20;

Best Local Similarity 100.0%; Pred. No. 3.3e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20

Db 1 TDFDQGWGPISYANGSGPDQ 20

RESULT 3

AAO18706

ID AAO18706 standard; Peptide; 20 AA.

XX AC AAO18706;

XX DT 24-OCT-2002 (first entry)

XX DE Hepatitis C virus E2 protein derived peptide E2-3B.

XX KW Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide;

XX KW immunostimulant; vaccine.

XX OS Hepatitis C virus.

XX PN WO200255548-A2.

XX PD 18-JUL-2002.

XX PF 11-JAN-2002; 2002WO-EP00219.

XX PR 11-JAN-2001; 2001US-260699P.

XX PR 30-AUG-2001; 2001US-315768P.

XX PA (INNO-) INNOGENETICS NV.

XX PI Maertens G, Bosman F, Buyse M;

XX DR WPI; 2002-599657/64.

XX PT New therapeutic vaccine compositions comprising at least one purified

PT recombinant hepatitis C virus (HCV) single or specific oligomeric

PT recombinant envelope protein E1 or E2, useful for immunizing humans

PT from HCV infection

XX PS Example 7; Page 227; 243pp; English.

XX CC The present invention relates to new therapeutic vaccine compositions for

CC inducing hepatitis C virus (HCV)-specific antibodies, comprising a

CC composition containing at least one purified recombinant HCV single or

CC specific oligomeric recombinant envelope proteins selected from an E1 and

CC an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are

CC useful for inducing HCV-specific antibodies or for immunising humans

CC against HCV. The recombinant HCV E1 and/or E2 proteins are useful as

CC vaccines or therapeutics, in HCV screening and confirmatory antibody

CC tests, for raising antibodies, in the preparation of medicament, and for

CC in vitro monitoring of HCV disease or prognosing the response to

CC treatment of patients suffering from HCV infection. The present sequence

CC is a peptide derived from the proteins of the invention.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 23; Length 20;

Best Local Similarity 100.0%; Pred. No. 3.3e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20

Db 1 TDFDQGWGPISYANGSGPDQ 20
|||||

RESULT 4

AAR13359 ID AAR13359 standard; Protein; 37 AA.

XX AC AAR13359;

XX DT 25-MAR-2003 (updated)

XX DT 23-OCT-1991 (first entry)

XX DE P447 HCV antigen (447-483).

XX KW C100-3; hepatitis C virus; immunoassay; epitope.

XX OS Synthetic.

XX FN AU9068390-A.

XX PD 27-JUN-1991.

XX PF 21-DEC-1990; 90AU-0068390.

XX PR 07-NOV-1990; 90US-0610180.

XX PR 22-DEC-1989; 89US-0456162.

XX PA (ABBO) ABBOTT LAB.

XX PA (LESN/) LESNIEWSKI R R.

XX DR WPI; 1991-238393/33.

XX PT Immunological assays for hepatitis C virus antibody - by using

XX PT polypeptide(s) contg. epitope(s) of hepatitis C virus antigens

XX PS Claim 1; Page 45; 62pp; English.

XX CC The polypeptide may be prepared by solid phase synthesis fragment

XX CC coupling (pref.) or using recombinant technology.

XX CC The assay has increased sensitivity and is more specific than

XX CC assays using the polypeptide C100-3 (EP-318216).

XX CC See also AAR13146-48 and AAR13343-65.

XX CC (Updated on 25-MAR-2003 to correct PA field.)

XX SQ Sequence 37 AA;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
|||||

Db 17 TDFDQGWGPISYANGSGPDQ 36

RESULT 5

AAR33878

ID AAR33878 standard; peptide; 37 AA.

XX AC AAR33878;

XX DT 25-MAR-2003 (updated)

XX DT 19-JUL-1993 (first entry)

XX DE Polypeptide p447 comprising HCV viral antigen.

XX KW Hepatitis C virus; NANBH; assay; antibody; p380-JH1; p380-J; p380LG;

XX KW p408.

XX OS Synthetic.

XX FN WO9306247-A1.

XX 01-APR-1993.
XX PD 16-SEP-1992; 92WO-US07813.
XX PF 16-SEP-1991; 91US-0760292.
XX PR (ABBO) ABBOTT LAB.
XX PA Lesniewski RR, Leung TK;
XX PI WPI; 1993-117563/14.
XX DR Assay for detecting presence of antibody to hepatitis C viral
XX PT antigen - by contacting sample with polypeptide contg. at least
XX PT one epitope of virus antigen
XX PS Disclosure; Page 14; 63pp; English.
XX CC The synthetic peptide p447 represents amino acid residues 447-483 of
XX CC the hepatitis C viral antigen. The peptide may be used in an assay to
XX CC detect antibodies to HCV and thus to diagnose chronic HCV infection.
XX CC See also AAR33861-87.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 37 AA;
XX Query Match 100.0%; Score 20; DB 14; Length 37;
XX Best Local Similarity 100.0%; Pred. No. 5.7e-14;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TDFDQGWGPISYANGSGPDQ 20
|||||
Db 17 TDFDQGWGPISYANGSGPDQ 36
RESULT 6
AAB18528
ID AAB18528 standard; Protein; 76 AA.
XX AC AAB18528;
XX DT 15-JAN-2001 (first entry)
XX DE Protein encoded by a novel hepatitis C virus cDNA clone 26j.
XX KW Hepatitis C virus; HCV; antisense polynucleotide; polypeptide;
XX KW viral infectivity; viral replication.
XX OS Hepatitis C virus.
XX PN EP1034785-A2.
XX PD 13-SEP-2000.
XX PF 16-MAR-1990; 2000EP-0109602.
XX PR 17-MAR-1989; 89US-0325338.
XX PR 20-APR-1989; 89US-0341334.
XX PR 18-MAY-1989; 89US-0355002.
XX PR 16-MAR-1990; 90EP-0302866.
XX (CHIR) CHIRON CORP.
XX PI Houghton M, Choo Q, Kuo G;
XX DR WPI; 2000-566891/53.
XX DR N-PSDB; AAA75284.
XX PT Novel composition comprising a hepatitis C virus antisense
XX PT polynucleotide which is complementary to or corresponds to a sense
XX PT strand of the virus genome, and selectively hybridises to it -

PS Example; Fig 4; 75pp; English.

XX The specification describes a pharmaceutical composition which
 CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
 CC HCV is characterized by a positive stranded RNA genome which has
 CC 40% homology at the polypeptide level to a HCV polypeptide. The
 CC antisense polynucleotide binds to cellular polynucleotides which
 CC enhance and/or are required for viral infectivity, replicative
 CC ability or chronicity. The antisense polynucleotides may also be
 CC designed to bind with high specificity, to be of increased stability,
 CC to be stable and to have low toxicity. The composition also comprises
 CC an agent which causes viral RNA to be inactive. The composition
 CC is used for preventing HCV replication in a system. The present
 CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
 CC course of the invention.

XX Sequence 76 AA;

Query Match 100.0%; Score 20; DB 21; Length 76;
 Best Local Similarity 100.0%; Pred. No. 1.1e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
 |||||
 Db 23 TDFDQGWGPISYANGSGPDQ 42

RESULT 7

AAB68043
 ID AAB68043 standard; protein; 254 AA.

AC AAB68043;

XX 29-JUN-2001 (first entry)

XX Amino acid sequence of water soluble variant of envelope E2 protein.

DE Eo protein; HCV; envelope protein; E2 protein; HCV infection;

XX HCV attachment.

XX Synthetic.

OS Hepatitis C virus.

XX WO200122984-A1.

XX 05-APR-2001.

XX 26-SEP-2000; 2000WO-US26395.

XX 29-SEP-1999; 99US-0407430.

XX (UYCO) UNIV COLUMBIA NEW YORK.

XX Worman HJ, Mamiya N;

XX WPI; 2001-273486/28.

XX Treating or preventing hepatitis C virus infection in a subject,
 PT involves administering hepatitis C virus envelope protein E2 binding
 PT agents -

XX Claim 5; Fig 8; 46pp; English.

XX The present sequence represents a water soluble variant of a Hepatitis C
 CC virus (HCV) envelope E2 protein. Agents that bind to the HCV E2 protein
 CC (such as the human Eo protein), and so inhibit the attachment of HCV onto
 CC cells (especially liver cells), are used to treat HCV infections in
 CC mammals, in particular humans. The specification also describes a method
 CC for identifying a compound which can be used for treating or preventing
 CC HCV in a subject and which can inhibit the attachment of HCV onto cells
 CC by inhibiting the binding of HCV envelope E2 protein to a cellular
 CC protein associated with HCV attachment and entry into cells. The method
 CC comprises incubating the compound, HCV envelope E2 protein or its variant

CC and a cellular protein capable of specifically binding to the HCV E2
 CC protein under suitable reaction conditions; determining the interactions
 CC between HCV envelope E2 protein and cellular protein in the presence and
 CC absence of the compound; and comparing the interaction to identify a
 CC compound which can inhibit the attachment of HCV onto cells.

XX Sequence 254 AA;

Query Match 100.0%; Score 20; DB 22; Length 254;

Best Local Similarity 100.0%; Pred. No. 3e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
 |||||
 Db 58 TDFDQGWGPISYANGSGPDQ 77

RESULT 8

AAR33997

ID AAR33997 standard; Protein; 278 AA.

XX AAR33997;

XX 25-MAR-2003 (updated)

DT 26-JUL-1993 (first entry)

XX Th E2/NS1 protein.

XX Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
 KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
 KW domain; immunological; cross-reactive; envelope protein; vaccine;
 KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.

XX Synthetic.

XX WO9306126-A1.

XX 01-APR-1993.

XX 11-SEP-1992; 92WO-US07683.

XX 13-SEP-1991; 91US-0759575.

XX (CHIR) CHIRON CORP.

XX Houghton M, Weiner AJ;

XX WPI; 1993-117468/14.

XX Immuno-reactive hepatitis C virus polypeptide compsns. - contg.
 at least 2 sequences from the first variable domain of distinct
 PT HCV isolates

XX Disclosure; Fig 3; 106pp; English.

XX The sequences given in AAR33992-002 represent a portion of the E2/NS1
 CC protein encoded by group I and group II HCV isolates, from amino acid
 CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
 CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
 CC 30 amino acids which shows large variation between nearly all isolates.
 CC This is an important immunoreactive domain. This putative envelope
 CC glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera
 CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
 CC flaviviruses, both of which confer protective immunity in hosts
 CC vaccinated with these polypeptides. It has been discovered that a
 CC number of important HCV epitopes vary among viral isolates and that
 CC these epitopes can be mapped to specific domains. This meant that
 CC immunologically cross-reactive polypeptides which focus on variable
 CC rather than constant domains can be produced. See also AAQ39134-48
 CC and AAR33982-91.

XX (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 278 AA;

```

Query Match      100.0%; Score 20; DB 14; Length 278;
Best Local Similarity 100.0%; Pred. No. 3.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFDQGWGPISYANGSGPDQ 20
   |||||
DB 94 TDFDQGWGPISYANGSGPDQ 113

RESULT 9
AAB68042
ID AAB68042 standard; protein; 363 AA.
XX AC AAB68042;
XX XX
XX 29-JUN-2001 (first entry)
XX DE Amino acid sequence of a Hepatitis C virus envelope E2 protein.
XX E2 protein; HCV; envelope protein; E2 protein; HCV infection;
XX HCV attachment.
XX OS Hepatitis C virus.
XX PN WO200122984-A1.
XX PD 05-APR-2001.
XX PF 26-SEP-2000; 2000WO-US26395.
XX PR 29-SEP-1999; 99US-0407430.
XX PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX PI Worman HJ, Mamiya N;
XX DR WPI; 2001-273486/28.
XX PT Treating or preventing hepatitis C virus infection in a subject.
XX PT involves administering hepatitis C virus envelope protein E2 binding
XX PT agents -
XX PS Claim 3; Fig 7; 46pp; English.
XX CC The present sequence represents a Hepatitis C virus (HCV) envelope
XX E2 protein. Agents that bind to the HCV E2 protein (such as the human
XX E2 protein), and so inhibit the attachment of HCV onto cells
XX (especially liver cells), are used to treat HCV infections in mammals,
XX in particular humans. The specification also describes a method for
XX identifying a compound which can be used for treating or preventing
XX HCV in a subject and which can inhibit the attachment of HCV onto cells
XX by inhibiting the binding of HCV envelope E2 protein to a cellular
XX protein associated with HCV attachment and entry into cells. The method
XX comprises incubating the compound, HCV envelope E2 protein or its variant
XX and a cellular protein capable of specifically binding to the HCV E2
XX protein under suitable reaction conditions; determining the interactions
XX between HCV envelope E2 protein and cellular protein in the presence and
XX absence of the compound; and comparing the interaction to identify a
XX compound which can inhibit the attachment of HCV onto cells.
XX SQ Sequence 363 AA;

Query Match      100.0%; Score 20; DB 22; Length 363;
Best Local Similarity 100.0%; Pred. No. 4.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFDQGWGPISYANGSGPDQ 20
   |||||
DB 80 TDFDQGWGPISYANGSGPDQ 99

RESULT 10

```

```

AAR34438
ID AAR34438 standard; Protein; 402 AA.
XX AC AAR34438;
XX XX
XX 25-MAR-2003 (updated)
XX DT 09-AUG-1993 (first entry)
XX XX
XX Sequence of glycoprotein E2/NS1 in clone HCV1.
XX KW Hepatitis C virus; envelope protein; glycoprotein; E2/NS1;
XX KW diagnostic reagent.
XX OS Hepatitis C virus.
XX PN EP537626-A1.
XX XX
XX 21-APR-1993.
XX PD
XX PF 08-OCT-1992; 92EP-0117191.
XX PR 08-OCT-1991; 91JP-0260824.
XX XX
XX (NAHE-) NAT INST OF HEALTH.
XX PI Harada S, Honda Y, Miyamura T, Saito I;
XX XX
XX WPI; 1993-127516/16.
XX DR N-PSDB; AAQ40330.
XX XX
XX Diagnostic reagent for hepatitis C virus - comprises second
XX envelope protein or first non-structural protein encoded by HCV
XX gene and has sugar chain
XX PS Claim 2; Pages 30-32; 58pp; English.
XX CC Glycoprotein E2/NS1 is derived from the second envelope protein or
XX first non-structural protein encoded by the genome of HCV. The
XX nucleic acid is extracted from the serum of the patient of hepatitis
XX C. The serum is pref. mixed with transfer RNA (tRNA) as a carrier
XX of virus RNA. As a technique of cloning cDNA from the nucleic acid,
XX it is preferred to use polymerase chain reaction method. In the
XX reaction, any commercially available random primers or synthesized
XX DNA having a base sequence similar to that of primer AS1 may be used
XX as a primer. Representative examples of sense primers include S1.
XX (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 402 AA;

Query Match      100.0%; Score 20; DB 14; Length 402;
Best Local Similarity 100.0%; Pred. No. 4.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFDQGWGPISYANGSGPDQ 20
   |||||
DB 124 TDFDQGWGPISYANGSGPDQ 143

RESULT 11
AAP90183
ID AAP90183 standard; protein; 454 AA.
XX AC AAP90183;
XX XX
XX 25-MAR-2003 (updated)
XX DT 01-NOV-1989 (first entry)
XX XX
XX Sequence of hepatitis C virus cDNA insert in clone k9-1.
XX DE
XX KW Hepatitis C virus; clone k9-1; probe; vaccine.
XX OS Pan troglodytes.
XX XX

```

```

FH Key          Location/Qualifiers
FT Region      97...454
FX
FX GB2212511-A.
FX
FX 26-JUL-1989.
FX
FX 18-NOV-1988; 88GB-0027024.
FX
FX 18-NOV-1987; 87US-0122714.
FX 30-DEC-1987; 87US-0139886.
FX 26-FEB-1988; 88US-0161072.
FX 26-OCT-1988; 88US-0263584.
FX
FX (CHIR ) CHIRON CORPORATION.
FX
FX Houghton M, Choo QL, Kuo G;
FX WPI; 1989-215054/30.
FX N-PSDB; AAN90335.
FX
FX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
FX polypeptide(s) and antibodies for diagnosis, prevention and
FX treatment of infection.
FX
FX Disclosure; fig 46; 235pp; English.
FX
FX The sequence is the peptide encoded by the hepatitis C virus
FX (HCV) cDNA insert in clone k9-1 (see AAN90335). The polypeptides
FX are used to diagnose HCV-induced NANBH, to raise antibodies for
FX immunosassay or treatment, or to produce vaccines.
FX The region shown overlaps the cDNA of AAN90327.
FX (Updated on 25-MAR-2003 to correct PR field.)
FX
SQ Sequence 454 AA;
SQ
Query Match          100.0%; Score 20; DB 10; Length 454;
Best Local Similarity 100.0%; Pred. No. 5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 TDFDQGWGPISYANGSGPDQ 20
   |||||
Db 13 TDFDQGWGPISYANGSGPDQ 32
   |||||

RESULT 12
AAP92049
ID AAP92049 standard; protein; 454 AA.
AC AAP92049;
XX
XX 25-MAR-2003 (updated)
DT 21-JAN-1991 (first entry)
XX
XX Sequence encoded by segment of the hepatitis C virus (HCV) cDNA sequence
XX in clone K9-1.
XX
XX Non-a non-B hepatitis; probe; vaccine; diagnosis;
KW passive immunotherapy; antigen.
XX
XX Hepatitis C virus.
OS
XX EP318216-A.
FN
XX 31-MAY-1989.
PD
XX 18-NOV-1988; 88EP-0310922.
FX
FX 18-NOV-1987; 87US-0122714.
FX 30-DEC-1987; 87US-0139886.
FX 26-FEB-1988; 88US-0161072.
FX 06-MAY-1988; 88US-0191263.
FX 26-OCT-1988; 88US-0263584.
FX

PR 14-NOV-1988; 88US-0271450.
XX (CHIR ) CHIRON CORP.
PA
XX Houghton M, Choo QL, Kuo G;
PI WPI; 1989-159274/22.
XX N-PSDB; AAN92105.
DR
XX Purified hepatitis C virus -
PT and assorted nucleic acids and polypeptide(s)
XX Example; Fig 46-1 - 46-2; 139pp; English.
XX
XX Purified hepatitis C virus (HCV) and purified or recombinant HCV nucleic
XX acids (NAs), encoding HCV polynucleotides or epitopes, and polypeptides
XX are claimed. HCV is a causative agent of non-A, non-B hepatitis (NANBH).
XX The NAs may be used to design probes for detn. of HCV NAs in samples.
XX The polypeptides may be used as immunoassay reagents and vaccines, and
XX to produce antibodies useful for diagnosis and passive immunotherapy.
XX The purified virus may also be used in vaccines.
XX (Updated on 25-MAR-2003 to correct PR field.)
CC (Updated on 25-MAR-2003 to correct PA field.)
CC
XX Sequence 454 AA;
SQ
Query Match          100.0%; Score 20; DB 10; Length 454;
Best Local Similarity 100.0%; Pred. No. 5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 TDFDQGWGPISYANGSGPDQ 20
   |||||
Db 13 TDFDQGWGPISYANGSGPDQ 32
   |||||

RESULT 13
AAB18526
ID AAB18526 standard; Protein; 454 AA.
XX
XX AAB18526;
AC
XX 15-JAN-2001 (first entry)
DT
XX Protein encoded by a novel hepatitis C virus cDNA clone k9-1.
DE
XX Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
KW viral infectivity; viral replication.
XX
XX Hepatitis C virus.
OS
XX EP1034785-A2.
FN
XX 13-SEP-2000.
PD
XX 16-MAR-1990; 2000EP-0109602.
FX
FX 17-MAR-1989; 89US-0325338.
FX 20-APR-1989; 89US-0341334.
FX 18-MAY-1989; 89US-0355002.
FX 16-MAR-1990; 90EP-0302866.
XX
XX (CHIR ) CHIRON CORP.
PA
XX Houghton M, Choo Q, Kuo G;
PI WPI; 2000-566891/53.
XX N-PSDB; AAA75282.
DR
XX Novel composition comprising a hepatitis C virus antisense
PT polynucleotide which is complementary to or corresponds to a sense
PT strand of the virus genome, and selectively hybridises to it -
XX Example; Fig 2; 75pp; English.
PS

```

XX The specification describes a pharmaceutical composition which
 CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
 CC HCV is characterized by a positive stranded RNA genome which has
 CC 40% homology at the polypeptide level to a HCV polypeptide. The
 CC antisense polynucleotide binds to cellular polynucleotides which
 CC enhance and/or are required for viral infectivity, replicative
 CC ability or chronicity. The antisense polynucleotides may also be
 CC designed to bind with high specificity, to be of increased stability,
 CC to be stable and to have low toxicity. The composition also comprises
 CC an agent which causes viral RNA to be inactive. The composition
 CC is used for preventing HCV replication in a system. The present
 CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
 CC course of the invention.

SQ Sequence 454 AA;
 Query Match 100.0%; Score 20; DB 21; Length 454;
 Best Local Similarity 100.0%; Pred. No. 5e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFDQGWGPISYANGSGPDQ 20
 |||||
 DB 13 TDFDQGWGPISYANGSGPDQ 32

RESULT 14
 AAR33992
 ID AAR33992 standard; Protein; 480 AA.

XX AAR33992;
 AC AAR33992;
 XX 25-MAR-2003 (updated)
 DT 26-JUL-1993 (first entry)
 XX HCV-1 E2/NS1 protein.
 XX Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
 KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
 KW domain; immunological; cross-reactive; envelope protein; vaccine;
 KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.

XX Synthetic.

XX WO9306126-A1.

XX 01-APR-1993.

XX 11-SEP-1992; 92WO-US07683.

XX 13-SEP-1991; 91US-0759575.

XX (CHIR) CHIRON CORP.

XX Houghton M, Weiner AJ;

XX WPI; 1993-117468/14.

XX Immuno-reactive hepatitis C virus polypeptide compens. - contg.
 PT at least 2 sequences from the first variable domain of distinct
 PT HCV isolates

XX Disclosure; Fig 3; 106pp; English.

XX The sequences given in AAR33992-002 represent a portion of the E2/NS1
 CC protein encoded by group I and group II HCV isolates, from amino acid
 CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
 CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
 CC 30 amino acids which shows large variation between nearly all isolates.
 CC This is an important immunoreactive domain. This putative envelope
 CC glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera
 CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
 CC flaviviruses, both of which confer protective immunity in hosts

CC vaccinated with these polypeptides. It has been discovered that a
 CC number of important HCV epitopes vary among viral isolates and that
 CC these epitopes can be mapped to specific domains. This meant that
 CC immunologically cross-reactive polypeptides which focus on variable
 CC rather than constant domains can be produced. See also AAQ39134-48
 CC and AAR33982-91.
 CC (Updated on 25-MAR-2003 to correct PN field.)

SQ Sequence 480 AA;

Query Match 100.0%; Score 20; DB 14; Length 480;
 Best Local Similarity 100.0%; Pred. No. 5.3e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFDQGWGPISYANGSGPDQ 20
 |||||
 DB 94 TDFDQGWGPISYANGSGPDQ 113

RESULT 15
 AAE02622

ID AAE02622 standard; Protein; 531 AA.

XX AAE02622;

XX 06-AUG-2001 (first entry)

XX Chimeric HCV E2661-HBsAg S domain encoded by pcMV-II-E2661-SAg.

XX Virus-like particle; immunogen; hepatitis B virus surface antigen;
 KW HBsAg; HBV; chimeric antigen; hepatitis C virus; HCV; virucide; vaccine;
 KW plasmid pcMV-II-E2661-SAg; E2 envelope glycoprotein; S domain.

XX Chimeric - Hepatitis C virus.

XX Chimeric - Hepatitis B virus.

XX Key Location/Qualifiers

FT Region 1..303

FT /note= "HCV 661 E2 envelope glycoprotein"

FT Region 306..531

FT /note= "HBsAg S domain"

XX WO200138358-A2.

XX 31-MAY-2001.

XX 22-NOV-2000; 2000WO-US32249.

XX 24-NOV-1999; 99US-0167224.

XX (CHIR) CHIRON CORP.

XX Selby M, Glazer E, Houghton M;

XX WPI; 2001-367661/38.

XX N-PSDB; AAD06793.

XX Virus-like particle for use as an immunogen, comprising a first
 PT hepatitis B virus surface antigen (HBsAg) and chimeric antigen
 PT comprising a second HBsAg covalently linked to hepatitis C immunogenic
 PT polypeptide -

XX Claim 29; Fig 4; 115pp; English.

XX The invention relates to a virus-like particle for use as an immunogen,
 CC comprising a first hepatitis B virus surface antigen (HBsAg) and a
 CC chimeric antigen comprising a second HBsAg which is covalently linked to
 CC an hepatitis C virus (HCV) immunogenic polypeptide, where the first and
 CC the second HBsAg each comprise a substantially complete S domain.
 CC The virus-like particle is useful as immunogen and as vaccine.
 CC The present sequence is a chimeric antigen comprising HCV 661 E2
 CC envelope glycoprotein and HBsAg S domain encoded by plasmid
 CC pcMV-II-E2661-SAg.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.6 Seconds
(without alignments)
79,832 Million cell updates/sec

Title: US-09-973-025-77
Perfect score: 20
Sequence: 1 TDFQGWGPISYANGSGPD 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-08-612-973-77
2	20	100.0	20	3	US-08-927-597-77
3	20	100.0	20	4	US-08-635-886C-33
4	20	100.0	37	5	PCT-US92-07813-18
5	20	100.0	45	1	US-08-262-037-35
6	20	100.0	50	1	US-08-262-037-39
7	20	100.0	52	1	US-08-262-037-36
8	20	100.0	57	1	US-08-262-037-40
9	20	100.0	76	3	US-08-444-818-79
10	20	100.0	278	1	US-08-440-103-15
11	20	100.0	278	1	US-08-440-542-15
12	20	100.0	278	1	US-08-231-368-15
13	20	100.0	278	1	US-08-440-210-15
14	20	100.0	278	4	US-09-046-604-15
15	20	100.0	402	1	US-08-460-806-13
16	20	100.0	402	1	US-08-325-630-13
17	20	100.0	402	2	US-08-483-695-39
18	20	100.0	403	2	US-07-965-285-39
19	20	100.0	403	2	US-08-487-231-39
20	20	100.0	403	3	US-09-201-912-39
21	20	100.0	454	3	US-08-444-818-73
22	20	100.0	480	1	US-08-440-103-14
23	20	100.0	480	1	US-08-440-542-14
24	20	100.0	480	1	US-08-231-368-14
25	20	100.0	480	1	US-08-440-210-14
26	20	100.0	480	4	US-09-046-604-14
27	20	100.0	663	3	US-08-824-057-3

28	20	100.0	663	4	US-09-415-582-3	Sequence 3, Appli
29	20	100.0	663	4	US-09-693-596-4	Sequence 4, Appli
30	20	100.0	2436	3	US-08-444-818-75	Sequence 75, Appl
31	20	100.0	2772	3	US-08-444-818-89	Sequence 89, Appl
32	20	100.0	2894	2	US-08-466-975A-23	Sequence 23, Appl
33	20	100.0	2894	2	US-08-331-671A-23	Sequence 23, Appl
34	20	100.0	2894	3	US-08-467-902A-23	Sequence 23, Appl
35	20	100.0	2894	3	US-09-275-265-23	Sequence 23, Appl
36	20	100.0	2894	4	US-09-941-611-23	Sequence 23, Appl
37	20	100.0	2955	2	US-08-443-260-3	Sequence 3, Appli
38	20	100.0	2955	3	US-08-442-805A-3	Sequence 3, Appli
39	20	100.0	2955	3	US-08-443-900A-3	Sequence 3, Appli
40	20	100.0	2955	3	US-08-444-818-124	Sequence 124, App
41	20	100.0	2955	3	US-08-249-843-3	Sequence 3, Appli
42	20	100.0	2955	3	US-08-444-818-138	Sequence 138, App
43	20	100.0	3011	1	US-08-440-103-36	Sequence 36, Appl
44	20	100.0	3011	1	US-08-440-542-36	Sequence 36, Appl
45	20	100.0	3011	1	US-07-910-760-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-612-973-77
; Sequence 77, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-612-973-77

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDFQGWGPISYANGSGPD 20
|||||

Db 1 TDFDQGWGPISYANGSGPDQ 20

RESULT 2

US-08-927-597-77
; Sequence 77, Application US/08927597
; Patent No. 6245503

; GENERAL INFORMATION:

; APPLICANT: MAERTENS, GEERT

; APPLICANT: BOSMAN, FONS

; APPLICANT: DE MARTYNOFF, GUY

; APPLICANT: BUYSE, MARIE-ANGE

; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

; NUMBER OF SEQUENCES: 111

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: NIXON & VANDERHYE P.C.

; STREET: 1100 NORTH GLEBE ROAD

; CITY: ARLINGTON

; STATE: VIRGINIA

; COUNTRY: U.S.A.

; ZIP: 22201-4714

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/927,597

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/612,973

; FILING DATE: 11-MAR-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: BYRNE, THOMAS E.

; REGISTRATION NUMBER: 32,205

; REFERENCE/DOCKET NUMBER: 1487-10

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 816-4000

; TELEFAX: (703) 816-4100

; INFORMATION FOR SEQ ID NO: 77:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 20 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-927-597-77

Query Match 100.0%; Score 20; DB 3; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.2e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20

Db 1 TDFDQGWGPISYANGSGPDQ 20

RESULT 3

US-08-635-886C-33

; Sequence 33, Application US/08635886C

; Patent No. 6555114

; GENERAL INFORMATION:

; APPLICANT: LEROUX-ROELS, Geert

; APPLICANT: DELEYS, Robert

; APPLICANT: MAERTENS, Geert

; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C

; FILE REFERENCE: 2752-18

; CURRENT APPLICATION NUMBER: US/08/635,886C

; CURRENT FILING DATE: 1996-04-25

; PRIOR APPLICATION NUMBER: PCT/EP94/03555

; PRIOR FILING DATE: 1994-10-28

; PRIOR APPLICATION NUMBER: EP 93402718.6

; PRIOR FILING DATE: 1993-11-04

; NUMBER OF SEQ ID NOS: 286

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 33

; LENGTH: 20

; TYPE: PRT

; ORGANISM: hepatitis C virus

US-08-635-886C-33

Query Match 100.0%; Score 20; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.2e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20

Db 1 TDFDQGWGPISYANGSGPDQ 20

RESULT 4

PCT-US92-07813-18

; Sequence 18, Application PC/TUS9207813

; GENERAL INFORMATION:

; APPLICANT: LESNIEWSKI, RICHARD R.

; APPLICANT: LEUNG, TAT K.

; TITLE OF INVENTION: HEPATITIS C ASSAY

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ABBOTT LABORATORIES CHAD377/AP6D

; STREET: ONE ABBOTT PARK ROAD

; CITY: ABBOTT PARK

; STATE: ILLINOIS

; COUNTRY: U.S.A.

; ZIP: 60064-3500

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/07813

; FILING DATE: 19920916

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: FOREMSKIP, PRISCILLA E.

; REGISTRATION NUMBER: 33,207

; REFERENCE/DOCKET NUMBER: 4767.P3.03

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 708-937-6365

; TELEFAX: 708-937-9556

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 37 amino acids

; TYPE: AMINO ACID

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

PCT-US92-07813-18

Query Match 100.0%; Score 20; DB 5; Length 37;

Best Local Similarity 100.0%; Pred. No. 2.1e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20

Db 17 TDFDQGWGPISYANGSGPDQ 36

RESULT 5

US-08-262-037-35

; Sequence 35, Application US/08262037

; Patent No. 5747239

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-262-037-36

Query Match 100.0%; Score 20; DB 1; Length 52;
Best Local Similarity 100.0%; Pred. No. 2.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
Db 12 TDFDQGWGPISYANGSGPDQ 31

RESULT 8
US-08-262-037-40
; Sequence 40, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990

ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 57 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-262-037-40

Query Match 100.0%; Score 20; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 3.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
Db 12 TDFDQGWGPISYANGSGPDQ 31

RESULT 9
US-08-444-818-79
; Sequence 79, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANOV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 76 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-444-818-79

Query Match 100.0%; Score 20; DB 3; Length 76;
Best Local Similarity 100.0%; Pred. No. 4.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20

Db 23 TDFDQGWGPISYANGSPDQ 42
|||||

RESULT 10
US-08-440-103-15
; Sequence 15, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440.103
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231.368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-103-15
Query Match 100.0%; Score 20; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TDFDQGWGPISYANGSPDQ 20
|||||
Db 94 TDFDQGWGPISYANGSPDQ 113
RESULT 11
US-08-440-542-15
; Sequence 15, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA

; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440.542
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231.368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-542-15
Query Match 100.0%; Score 20; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TDFDQGWGPISYANGSPDQ 20
|||||
Db 94 TDFDQGWGPISYANGSPDQ 113
RESULT 12
US-08-231-368-15
; Sequence 15, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231.368
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-231-368-15
Query Match 100.0%; Score 20; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TDFDQGWGPISYANGSPDQ 20
|||||
Db 94 TDFDQGWGPISYANGSPDQ 113

TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 278 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-231-368-15

Query Match 100.0%; Score 20; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
|||||
Db 94 TDFDQGWGPISYANGSGPDQ 113

RESULT 13
US-08-440-210-15
Sequence 15, Application US/08440210
Patent No. 5766845
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,210
FILING DATE: 12-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 278 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-210-15

Query Match 100.0%; Score 20; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
|||||

Db 94 TDFDQGWGPISYANGSGPDQ 113

RESULT 14
US-09-046-604-15
Sequence 15, Application US/09046604
Patent No. 6303292
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/046,604
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 278 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-046-604-15

Query Match 100.0%; Score 20; DB 4; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
|||||
Db 94 TDFDQGWGPISYANGSGPDQ 113

RESULT 15
US-08-460-806-13
Sequence 13, Application US/08460806
Patent No. 5747241
GENERAL INFORMATION:
APPLICANT: MIYAMURA, TATSUO
APPLICANT: SAITO, IZUMU
APPLICANT: HARADA, SHIZUKO
APPLICANT: HONDA, YOSHIKAZU
TITLE OF INVENTION: DIAGNOSTIC REAGENT FOR HEPATITIS C
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCQUELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400

CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460.806
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/325.630
FILING DATE: 19-OCT-1994
APPLICATION NUMBER: US 07/956.993
FILING DATE: 06-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5747241man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 4667-001-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 402 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-460-806-13

Query Match 100.0%; Score 20; DB 1; Length 402;
Best Local Similarity 100.0%; Pred. No. 1.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
| | | | | | | | | | | | | | | | | | | | | |
Db 124 TDFDQGWGPISYANGSGPDQ 143

Search completed: November 21, 2003, 21:15:15
Job time : 10.6 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time 19.5 Seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-77

Perfect score: 20

Sequence: 1 TDFDQGWGPISYANGSGPDQ 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	10	US-09-973-025-77
2	20	100.0	20	11	US-09-899-303-77
3	20	100.0	20	11	US-09-995-808-77
4	20	100.0	20	11	US-09-995-860-77
5	20	100.0	20	12	US-09-995-791-77
6	20	100.0	254	10	US-09-407-430-3
7	20	100.0	363	10	US-09-407-430-2
8	20	100.0	637	12	US-10-187-257-4
9	20	100.0	637	12	US-10-265-083-2
10	20	100.0	2894	10	US-09-941-611-23
11	20	100.0	2894	15	US-10-044-995-23
12	20	100.0	3011	9	US-09-916-359-2
13	20	100.0	3011	16	US-10-232-643-6
14	13	65.0	250	10	US-09-952-572-8
15	13	65.0	350	10	US-09-929-955-4

16	13	65.0	350	14	US-10-104-966-4	Sequence 4, Appli
17	13	65.0	363	12	US-10-128-587A-97	Sequence 97, Appl
18	13	65.0	363	15	US-10-128-590-97	Sequence 97, Appl
19	13	65.0	3011	9	US-09-742-659-4	Sequence 4, Appli
20	13	65.0	3011	10	US-09-238-076-20	Sequence 20, Appl
21	13	65.0	3011	10	US-09-522-572-9	Sequence 9, Appli
22	13	65.0	3011	10	US-09-929-955-1	Sequence 1, Appli
23	13	65.0	3011	10	US-09-747-419-20	Sequence 20, Appl
24	13	65.0	3011	11	US-09-891-894-3	Sequence 3, Appli
25	13	65.0	3011	11	US-09-995-937-20	Sequence 20, Appl
26	13	65.0	3011	11	US-09-317-563-20	Sequence 20, Appl
27	13	65.0	3011	12	US-10-184-150-3	Sequence 3, Appli
28	13	65.0	3011	14	US-10-104-966-1	Sequence 1, Appli
29	13	65.0	3011	15	US-10-259-275-20	Sequence 20, Appl
30	13	65.0	3012	10	US-09-238-076-2	Sequence 2, Appli
31	13	65.0	3012	11	US-09-995-937-2	Sequence 2, Appli
32	13	65.0	3012	11	US-09-317-563-2	Sequence 2, Appli
33	11	55.0	176	10	US-09-321-397-81	Sequence 81, Appl
34	9	45.0	14	11	US-09-899-046-267	Sequence 267, App
35	9	45.0	14	11	US-09-878-281-267	Sequence 267, App
36	9	45.0	37	10	US-09-921-397-36	Sequence 36, Appl
37	8	40.0	20	10	US-09-973-025-76	Sequence 76, Appl
38	8	40.0	20	10	US-09-973-025-78	Sequence 78, Appl
39	8	40.0	20	11	US-09-899-303-76	Sequence 76, Appl
40	8	40.0	20	11	US-09-899-303-78	Sequence 78, Appl
41	8	40.0	20	11	US-09-995-808-76	Sequence 76, Appl
42	8	40.0	20	11	US-09-995-808-78	Sequence 78, Appl
43	8	40.0	20	11	US-09-995-860-76	Sequence 76, Appl
44	8	40.0	20	11	US-09-995-860-78	Sequence 78, Appl
45	8	40.0	20	12	US-09-995-791-76	Sequence 76, Appl

ALIGNMENTS

RESULT 1

US-09-973-025-77
; Sequence 77, Application US/09973025
; Publication No. US20020182706A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100

1 TDFDQGWGPISYANGSGPDQ

; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-77

Query Match 100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
|||||
Db 1 TDFDQGWGPISYANGSGPDQ 20
|||||

RESULT 6

US-09-407-430-3
; Sequence 3, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto

; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; TITLE OF INVENTION: HEPATITIS C VIRUS INFECTION
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-407-430-3

Query Match 100.0%; Score 20; DB 10; Length 254;
Best Local Similarity 100.0%; Pred. No. 5.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
|||||
Db 58 TDFDQGWGPISYANGSGPDQ 77
|||||

RESULT 7

US-09-407-430-2
; Sequence 2, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto

; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; TITLE OF INVENTION: HEPATITIS C VIRUS INFECTION
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-09-407-430-2

Query Match 100.0%; Score 20; DB 10; Length 363;
Best Local Similarity 100.0%; Pred. No. 7.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
|||||
Db 80 TDFDQGWGPISYANGSGPDQ 99
|||||

RESULT 8

US-10-187-257-4
; Sequence 4, Application US/10187257
; Publication No. US20030138458A1
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: CORTES, Steve

; TITLE OF INVENTION: HCV E1E2 VACCINE COMPOSITIONS
; FILE REFERENCE: 2302-17206
; CURRENT APPLICATION NUMBER: US/10/187,257
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
; OTHER INFORMATION: amino acid
US-10-187-257-4

Query Match 100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 1.1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
|||||
Db 291 TDFDQGWGPISYANGSGPDQ 310
|||||

RESULT 9

US-10-265-083-2
; Sequence 2, Application US/10265083
; Publication No. US20030170273A1
; GENERAL INFORMATION:
; APPLICANT: O'HAGAN, Derek
; APPLICANT: VALIANTE, Nicholas

; TITLE OF INVENTION: ADJUVANT COMPOSITIONS
; FILE REFERENCE: 2302-18357.30
; CURRENT APPLICATION NUMBER: US/10/265,083
; CURRENT FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7
; OTHER INFORMATION: region
US-10-265-083-2

Query Match 100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 1.1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
|||||
Db 291 TDFDQGWGPISYANGSGPDQ 310
|||||

RESULT 10

US-09-941-611-23
; Sequence 23, Application US/09941611
; Patent No. US20020106640A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSWIN, HUGO

; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; ANTIBODIES TO HEPATITIS C VIRUS

```
;
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/941,611
; FILING DATE: 30-Aug-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/391,671
; FILING DATE: 1995-02-21
; APPLICATION NUMBER: WO PCT/EP91/02409
; FILING DATE: 13-DEC-1991
; APPLICATION NUMBER: EP 90124241.2
; FILING DATE: 14-DEC-1990
;
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 1487-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 7038164000
; TELEFAX: 7038164100
;
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-941-611-23

Query Match 100.0%; Score 20; DB 10; Length 2894;
Best Local Similarity 100.0%; Pred. No. 4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFDQGWGPISYANGSGPDQ 20
Db 463 TDFDQGWGPISYANGSGPDQ 482

RESULT 11
US-10-044-995-23
; Sequence 23, Application US/10044995
; Publication No. US20030049685A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; POLLET, DIRK
; MAERTENS, GEERT
; VAN HEUVERSWUN, HUGO
;
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; MEDIUM TYPE: Floppy disk
```

```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/044,995
; FILING DATE: 15-Jan-2002
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/391,671
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 07/920,286
; FILING DATE: 14-OCT-1992
; APPLICATION NUMBER: WO PCT/EP91/02409
; FILING DATE: 13-DEC-1991
; APPLICATION NUMBER: EP 90124241.2
; FILING DATE: 14-DEC-1990
;
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 1487-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 7038164000
; TELEFAX: 7038164100
;
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-044-995-23

Query Match 100.0%; Score 20; DB 15; Length 2894;
Best Local Similarity 100.0%; Pred. No. 4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFDQGWGPISYANGSGPDQ 20
Db 463 TDFDQGWGPISYANGSGPDQ 482

RESULT 12
US-09-916-359-2
; Sequence 2, Application US/09916359
; Patent No. US20020034734A1
; GENERAL INFORMATION:
; APPLICANT: Veronique Barban
; TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
; TITLE OF INVENTION: TREATING C HEPATITIS
; FILE REFERENCE: PWC97-03A
; CURRENT APPLICATION NUMBER: US/09/916,359
; CURRENT FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 09/388,874
; PRIOR FILING DATE: 1999-09-02
; PRIOR APPLICATION NUMBER: 97/02,887
; PRIOR FILING DATE: 1997-03-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Virus
; US-09-916-359-2

Query Match 100.0%; Score 20; DB 9; Length 3011;
Best Local Similarity 100.0%; Pred. No. 4.1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFDQGWGPISYANGSGPDQ 20
Db 463 TDFDQGWGPISYANGSGPDQ 482
```

Db 463 TDFQGWGPISYANGSPDQ 482

RESULT 13

US-10-232-643-6

Sequence 6, Application US/10232643

Publication No. US20030129586A1

GENERAL INFORMATION:

APPLICANT: HOUGHTON, MICHAEL

CHOO, QUI-LIM

HAN, JANG

CHOE, JOONHO

TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING

HELICASE ACTIVITY AND IMPROVED SOLUBILITY

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: CHIRON CORPORATION

STREET: Intellectual Property - R440, P.O. Box 8097

CITY: Emeryville

STATE: California

COUNTRY: U.S.A.

ZIP: 94662-8097

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/232,643

FILING DATE: 30-Aug-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/483,799

FILING DATE: <Unknown>

APPLICATION NUMBER: 08/529,169

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Harbin, Alisa A.

REGISTRATION NUMBER: 33,895

REFERENCE/DOCKET NUMBER: 0100.005

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 923-3274

TELEFAX: (510) 655-3542

TELEX: n/a

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 3011 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Duplication

LOCATION: 9

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Lys or Arg"

FEATURE:

NAME/KEY: Duplication

LOCATION: 11

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Asn or Thr"

FEATURE:

NAME/KEY: Duplication

LOCATION: 176

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Ile or Thr"

FEATURE:

NAME/KEY: Duplication

LOCATION: 334

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Met or Val"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2502

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Leu or Phe"

LOCATION: 603

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Leu or Ile"

FEATURE:

NAME/KEY: Duplication

LOCATION: 848

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Tyr or Asn"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1114

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Pro or Ser"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1117

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Ser or Thr"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1276

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Pro or Leu"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1454

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Cys or Tyr"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1471

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Thr or Ser"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1877

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Glu or Gly"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1948

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Leu or His"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1949

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Ser or Cys"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2021

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Gly or Val"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2349

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Thr or Ser"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2385

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Tyr or Phe"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2386

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Ser or Ala"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2502

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Leu or Phe"

```
;
;
; NAME/KEY: Duplication
; LOCATION: 2690
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Arg or Gly"
;
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2921
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Arg or Gly"
;
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2996
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Leu or Pro"
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-232-643-6

Query Match      100.0%; Score 20; DB 16; Length 3011;
Best Local Similarity 100.0%; Pred. No. 4.1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
   |||||
Db 463 TDFDQGWGPISYANGSGPDQ 482

RESULT 14
US-09-952-572-8
; Sequence 8, Application US/09952572
; Patent No. US20020119495A1
; GENERAL INFORMATION:
; APPLICANT: HAWAII BIOTECHNOLOGY GROUP, Inc.
; APPLICANT: NAKANO, Eileen
; APPLICANT: CLEMENTS, David
; APPLICANT: HUMPHREYS, Tom
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITION OF HEPATITIS C
; FILE REFERENCE: HAWBIO1100
; CURRENT APPLICATION NUMBER: US/09/952,572
; CURRENT FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: US 60/230,927
; PRIOR FILING DATE: 2000-09-13
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-09-952-572-8

Query Match      65.0%; Score 13; DB 10; Length 250;
Best Local Similarity 100.0%; Pred. No. 6.9e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 QGWGPISYANGSG 17
   |||||
Db 56 QGWGPISYANGSG 68

RESULT 15
US-09-929-955-4
; Sequence 4, Application US/09929955
; Patent No. US20020136740A1
; GENERAL INFORMATION:
; APPLICANT: Matti Salberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; FILE REFERENCE: TRIPEP.23AUS2
; CURRENT APPLICATION NUMBER: US/09/929,955
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03

;
;
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/225,767
; PRIOR FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE: 0
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-09-929-955-4

Query Match      65.0%; Score 13; DB 10; Length 350;
Best Local Similarity 100.0%; Pred. No. 9.1e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 QGWGPISYANGSG 17
   |||||
Db 88 QGWGPISYANGSG 100

Search completed: November 21, 2003, 22:19:38
Job time : 20.5 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 163.1 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-77

Perfect score: 20

Sequence: 1 TDFQGMGPTSYANGSGPDQ 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA_Main.*

- 1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/1/paa/US097A_COMB.pep.*
- 22: /cgn2_6/ptodata/1/paa/US097B_COMB.pep.*
- 23: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
- 24: /cgn2_6/ptodata/1/paa/US099A_COMB.pep.*
- 25: /cgn2_6/ptodata/1/paa/US099B_COMB.pep.*
- 26: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
- 27: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
- 28: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
- 29: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
- 30: /cgn2_6/ptodata/1/paa/US104_COMB.pep.*
- 31: /cgn2_6/ptodata/1/paa/US106_COMB.pep.*
- 32: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	20	100.0	20	13	US-08-974-685-33
					Sequence 33, Appl

2	20	100.0	20	13	US-08-974-690-33	Sequence 33, Appl
3	20	100.0	20	13	US-08-974-690A-33	Sequence 33, Appl
4	20	100.0	20	13	US-08-974-690B-33	Sequence 33, Appl
5	20	100.0	20	13	US-08-974-690C-33	Sequence 33, Appl
6	20	100.0	20	23	US-09-899-303-77	Sequence 77, Appl
7	20	100.0	20	23	US-09-899-303A-77	Sequence 77, Appl
8	20	100.0	20	25	US-09-973-025-77	Sequence 77, Appl
9	20	100.0	20	25	US-09-995-791-77	Sequence 77, Appl
10	20	100.0	20	25	US-09-995-808-77	Sequence 77, Appl
11	20	100.0	20	25	US-09-995-860-77	Sequence 77, Appl
12	20	100.0	20	26	US-10-020-510-77	Sequence 77, Appl
13	20	100.0	20	29	US-10-321-798-77	Sequence 77, Appl
14	20	100.0	27	3	US-07-748-564A-129	Sequence 129, App
15	20	100.0	27	3	US-07-819-360A-53	Sequence 53, Appl
16	20	100.0	37	9	US-08-507-740-18	Sequence 18, Appl
17	20	100.0	37	13	US-08-905-054-18	Sequence 18, Appl
18	20	100.0	37	13	US-08-905-054B-18	Sequence 18, Appl
19	20	100.0	45	8	US-08-475-482-35	Sequence 35, Appl
20	20	100.0	45	8	US-08-477-072-35	Sequence 35, Appl
21	20	100.0	45	8	US-08-477-582-35	Sequence 35, Appl
22	20	100.0	45	8	US-08-480-253-35	Sequence 35, Appl
23	20	100.0	50	8	US-08-475-482-39	Sequence 39, Appl
24	20	100.0	50	8	US-08-477-072-39	Sequence 39, Appl
25	20	100.0	50	8	US-08-477-582-39	Sequence 39, Appl
26	20	100.0	50	8	US-08-480-253-39	Sequence 39, Appl
27	20	100.0	52	8	US-08-475-482-36	Sequence 36, Appl
28	20	100.0	52	8	US-08-477-072-36	Sequence 36, Appl
29	20	100.0	52	8	US-08-477-582-36	Sequence 36, Appl
30	20	100.0	52	8	US-08-480-253-36	Sequence 36, Appl
31	20	100.0	57	8	US-08-475-482-40	Sequence 40, Appl
32	20	100.0	57	8	US-08-477-072-40	Sequence 40, Appl
33	20	100.0	57	8	US-08-477-582-40	Sequence 40, Appl
34	20	100.0	57	8	US-08-480-253-40	Sequence 40, Appl
35	20	100.0	76	8	US-08-403-590B-79	Sequence 79, Appl
36	20	100.0	76	8	US-08-444-112-79	Sequence 79, Appl
37	20	100.0	122	8	US-08-436-966-5	Sequence 5, Appl
38	20	100.0	254	1	PCT-US00-26395-3	Sequence 3, Appl
39	20	100.0	254	18	US-09-407-430-3	Sequence 15, Appl
40	20	100.0	278	8	US-08-471-498-15	Sequence 3, Appl
41	20	100.0	333	30	US-10-445-724-3	Sequence 4, Appl
42	20	100.0	333	32	US-60-409-909-4	Sequence 2, Appl
43	20	100.0	363	1	PCT-US00-26395-2	Sequence 2, Appl
44	20	100.0	363	18	US-09-407-430-2	Sequence 9, Appl
45	20	100.0	416	8	US-08-436-966-9	

ALIGNMENTS

RESULT 1
US-08-974-685-33
; Sequence 33, Application US/08974685
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; DELEYS, ROBERT
; MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
; THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 181
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,685
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 33:
US-08-974-685-33

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
Db 1 TDFDQGWGPISYANGSGPDQ 20

RESULT 2

US-08-974-690-33
Sequence 33, Application US/08974690

GENERAL INFORMATION:

APPLICANT: LEROUIX-ROELS, GEERT

APPLICANT: DELEYS, ROBERT

APPLICANT: MAERTENS, GEERT

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF

TITLE OF INVENTION: HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 166

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,690

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/635,886

FILING DATE: 25-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-11

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-974-690-33

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
Db 1 TDFDQGWGPISYANGSGPDQ 20

RESULT 3

US-08-974-690A-33

Sequence 33, Application US/08974690A

GENERAL INFORMATION:

APPLICANT: LEROUIX-ROELS, GEERT

APPLICANT: DELEYS, ROBERT

APPLICANT: MAERTENS, GEERT

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF

TITLE OF INVENTION: HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 177

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,690A

FILING DATE: 19-Nov-1997

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: SADOFF, B. J.

REGISTRATION NUMBER: 36,663

REFERENCE/DOCKET NUMBER: 2752-20

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 33:

US-08-974-690A-33

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
Db 1 TDFDQGWGPISYANGSGPDQ 20

RESULT 4

US-08-974-690B-33

Sequence 33, Application US/08974690B

GENERAL INFORMATION:

APPLICANT: LEROUIX-ROELS, GEERT

APPLICANT: DELEYS, ROBERT

APPLICANT: MAERTENS, GEERT

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF

TITLE OF INVENTION: HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690B
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 33:
US-08-974-690B-33

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
Db 1 TDFDQGWGPISYANGSGPDQ 20

RESULT 5
US-08-974-690C-33
Sequence 33, Application US/08974690C
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELEYS, Robert
APPLICANT: MAERTENS, Geert
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
FILE REFERENCE: 2551-94
CURRENT APPLICATION NUMBER: US/08/974,690C
CURRENT FILING DATE: 1997-11-19
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28
PRIOR APPLICATION NUMBER: EP 93402718.6
PRIOR FILING DATE: 1993-11-04
NUMBER OF SEQ ID NOS: 286
SOFTWARE: Patent In version 3.1
SEQ ID NO 33
LENGTH: 20
TYPE: PRT
ORGANISM: hepatitis C virus
US-08-974-690C-33

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20

Db 1 TDFDQGWGPISYANGSGPDQ 20

RESULT 6
US-09-899-303-77
Sequence 77, Application US/09899303
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-09-899-303-77

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20

Db 1 TDFDQGWGPISYANGSGPDQ 20

RESULT 7
US-09-899-303A-77
Sequence 77, Application US/09899303A
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.

STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303A
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-09-899-303A-77
Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TDFDQGWGPISYANGSGPDQ 20
Db 1 TDFDQGWGPISYANGSGPDQ 20
RESULT 8
US-09-973-025-77
; Sequence 77, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/973,025
FILING DATE: 10-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-09-973-025-77
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TDFDQGWGPISYANGSGPDQ 20
Db 1 TDFDQGWGPISYANGSGPDQ 20
RESULT 9
US-09-995-791-77
; Sequence 77, Application US/09995791
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 77
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-77
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TDFDQGWGPISYANGSGPDQ 20
Db 1 TDFDQGWGPISYANGSGPDQ 20
RESULT 10
US-09-995-808-77
; Sequence 77, Application US/09995808
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 77
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-77
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDFDQGWGPISYANGSGPDQ 20
| | | | | | | | | | | | | | | | | | | | | |
DB 1 TDFDQGWGPISYANGSGPDQ 20
| | | | | | | | | | | | | | | | | | | | | |

RESULT 11
US-09-995-860-77
; Sequence 77, Application US/09995860
; GENERAL INFORMATION:
; APPLICANT: Immunogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 77
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-77

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDFDQGWGPISYANGSGPDQ 20
| | | | | | | | | | | | | | | | | | | | | |
DB 1 TDFDQGWGPISYANGSGPDQ 20
| | | | | | | | | | | | | | | | | | | | | |

RESULT 12
US-10-020-510-77
; Sequence 77, Application US/10020510
; GENERAL INFORMATION:
; APPLICANT: Immunogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-72
; CURRENT APPLICATION NUMBER: US/10/020,510
; CURRENT FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 77
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-020-510-77

Query Match 100.0%; Score 20; DB 26; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDFDQGWGPISYANGSGPDQ 20
| | | | | | | | | | | | | | | | | | | | | |
DB 1 TDFDQGWGPISYANGSGPDQ 20
| | | | | | | | | | | | | | | | | | | | | |

RESULT 13
US-10-321-798-77
; Sequence 77, Application US/10321798
; GENERAL INFORMATION:
; APPLICANT: Immunogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-93
; CURRENT APPLICATION NUMBER: US/10/321,798
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: 60/418,358
; PRIOR FILING DATE: 2002-10-16
```

```
; PRIOR APPLICATION NUMBER: 10/020,510
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 77
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-321-798-77

Query Match 100.0%; Score 20; DB 29; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDFDQGWGPISYANGSGPDQ 20
| | | | | | | | | | | | | | | | | | | | | |
DB 1 TDFDQGWGPISYANGSGPDQ 20
| | | | | | | | | | | | | | | | | | | | | |

RESULT 14
US-07-748-564A-129
; Sequence 129, Application US/07748564A
; GENERAL INFORMATION:
; APPLICANT: Zebedee, Suzanne
; APPLICANT: Inchauspe, Genevieve
; APPLICANT: Nasoff, Marc
; APPLICANT: Prince, Alfred
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS ANTIGEN,
; TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 137
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DRESSLER, GOLDSMITH, SHORE, SUTKER &
; ADDRESSEE: MILNAMOW, LTD.
; STREET: 180 N. Stetson, Suite 4700
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/748,564A
; FILING DATE: 19910821
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/61369
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/573643
; FILING DATE: 25-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gamson, Edward P.
; REGISTRATION NUMBER: 29,381
; REFERENCE/DOCKET NUMBER: PHA0029P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-616-5400
; TELEFAX: 312-616-5460
; INFORMATION FOR SEQ ID NO: 129:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-748-564A-129

Query Match 100.0%; Score 20; DB 3; Length 27;
Best Local Similarity 100.0%; Pred. No. 2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDFDQGWGPISYANGSGPDQ 20
```

Db 8 TDFDQGWGPISYANGSPDQ 27
|||||

RESULT 15

US-07-819-360A-53
; Sequence 53, Application US/07819360A
; GENERAL INFORMATION:
; APPLICANT: Prince, Alfred
; APPLICANT: Ke, Liu
; APPLICANT: Inchauspe, Genevieve
; TITLE OF INVENTION: A Chinese Isolate of Non-A, Non-B
; TITLE OF INVENTION: Hepatitis Virus: Compositions and Methods
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dressler, Goldsmith, Shore, Sutker & Milnamow
; STREET: 180 N. Stetson St.
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/819,360A
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Northrup, Thomas E
; REGISTRATION NUMBER: 33,268
; REFERENCE/DOCKET NUMBER: hydoooolp
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 13126165400
; TELEFAX: 13126165460
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-819-360A-53

Query Match 100.0%; Score 20; DB 3; Length 27;
Best Local Similarity 100.0%; Pred. No. 2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSPDQ 20
Db 8 TDFDQGWGPISYANGSPDQ 27
|||||

Search completed: November 21, 2003, 22:09:51
Job time : 163.1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.55 Seconds
(without alignments)
117.536 Million cell updates/sec

Title: US-09-973-025-77

Perfect score: 20

Sequence: 1 TDFDQGMGPISYANGSGPDQ 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents AA_Nev.*

- 1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	6	US-10-651-165-33
2	20	100.0	333	1	PCT-US03-19834-3
3	20	100.0	333	6	US-10-655-562-4
4	20	100.0	637	1	PCT-US03-33610-4
5	20	100.0	3011	1	PCT-US03-19834-2
6	20	100.0	3011	6	US-10-296-734-406
7	19	95.0	30	6	US-10-296-734-470
8	19	95.0	2010	6	US-10-296-734-814
9	19	95.0	5985	6	US-10-296-734-810
10	16	80.0	30	6	US-10-296-734-468
11	16	80.0	1997	6	US-10-296-734-816
12	9	45.0	14	5	US-09-638-693-267
13	9	45.0	14	5	US-09-873-224A-267
14	8	40.0	20	6	US-10-651-165-32
15	8	40.0	20	6	US-10-651-165-34
16	8	40.0	1026	1	PCT-US03-20409-3
17	6	30.0	136	1	PCT-US03-23932-4
18	6	30.0	176	5	US-09-638-693-158
19	6	30.0	176	5	US-09-873-224A-158
20	6	30.0	238	6	US-10-679-063-17338
21	6	30.0	247	6	US-10-425-114A-43167
22	6	30.0	322	6	US-10-425-114A-44496
23	6	30.0	336	6	US-10-679-063-2311
24	6	30.0	347	6	US-10-664-391-9
25	6	30.0	390	6	US-10-679-063-2533
26	6	30.0	390	6	US-10-679-063-2534

27	6	30.0	539	6	US-10-664-391-11	Sequence 11, Appl
28	6	30.0	650	1	PCT-US03-14382-79	Sequence 79, Appl
29	6	30.0	650	1	PCT-US03-14382-80	Sequence 80, Appl
30	6	30.0	650	1	PCT-US03-14382-81	Sequence 81, Appl
31	6	30.0	650	1	PCT-US03-14382-82	Sequence 82, Appl
32	6	30.0	650	1	PCT-US03-14382-83	Sequence 83, Appl
33	6	30.0	650	1	PCT-US03-14382-84	Sequence 84, Appl
34	6	30.0	650	1	PCT-US03-14382A-79	Sequence 79, Appl
35	6	30.0	650	1	PCT-US03-14382A-80	Sequence 80, Appl
36	6	30.0	650	1	PCT-US03-14382A-81	Sequence 81, Appl
37	6	30.0	650	1	PCT-US03-14382A-82	Sequence 82, Appl
38	6	30.0	650	1	PCT-US03-14382A-83	Sequence 83, Appl
39	6	30.0	650	1	PCT-US03-14382A-84	Sequence 84, Appl
40	6	30.0	771	6	US-10-679-063-2598	Sequence 211, App
41	6	30.0	2280	1	PCT-US03-20322-211	Sequence 211, App
42	5	25.0	8	5	US-09-641-528B-10146	Sequence 10146, A
43	5	25.0	8	5	US-09-641-528B-34433	Sequence 34433, A
44	5	25.0	8	5	US-09-641-528B-41750	Sequence 41750, A
45	5	25.0	9	5	US-09-641-528B-10114	Sequence 10114, A

ALIGNMENTS

RESULT 1
US-10-651-165-33
; Sequence 33, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-33

Query Match 100.0% Score 20; DB 6; Length 20;
Best Local Similarity 100.0%; Pred.No. 1.7e-09; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

QY 1 TDFDQGMGPISYANGSGPDQ 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 TDFDQGMGPISYANGSGPDQ 20

RESULT 2
PCT-US03-19834-3
; Sequence 3, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; TITLE OF INVENTION: METHODS FOR TREATING HYPERLIPIDEMIA
; FILE REFERENCE: IOWA:045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-05-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158

```
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-3

Query Match      100.0%; Score 20; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 5.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TDFDQGWGPISYANGSGPDQ 20
      |||||
Db      81 TDFDQGWGPISYANGSGPDQ 100

RESULT 3
US-10-655-562-4
; Sequence 4, Application US/10655562
; GENERAL INFORMATION:
; APPLICANT: BISHOP-HURLEY, SHARON L.
; APPLICANT: SCHMIDT, FRANCIS J.
; APPLICANT: SMITH, ARNOLD L.
; TITLE OF INVENTION: PHAGE-DISPLAY PEPTIDES AS NOVEL ANTIMICROBIAL AGENTS
; TITLE OF INVENTION: AGAINST HAEMOPHILUS INFLUENZAE
; FILE REFERENCE: UWO:022US
; CURRENT APPLICATION NUMBER: US/10/655,562
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: 60,409,909
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-10-655-562-4

Query Match      100.0%; Score 20; DB 6; Length 333;
Best Local Similarity 100.0%; Pred. No. 5.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TDFDQGWGPISYANGSGPDQ 20
      |||||
Db      81 TDFDQGWGPISYANGSGPDQ 100

RESULT 4
PCT-US03-33610-4
; Sequence 4, Application PC/TUS0333610
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: SELBY, Mark
; TITLE OF INVENTION: ACTIVATION OF HCV-SPECIFIC T-CELLS
; FILE REFERENCE: 2300-1612.60
; CURRENT APPLICATION NUMBER: PCT/US03/33610
; CURRENT FILING DATE: 2003-10-24
; PRIOR APPLICATION NUMBER: 10/281,341
; PRIOR FILING DATE: 2002-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
PCT-US03-33610-4

Query Match      100.0%; Score 20; DB 1; Length 637;
Best Local Similarity 100.0%; Pred. No. 7.4e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TDFDQGWGPISYANGSGPDQ 20
      |||||
Db      291 TDFDQGWGPISYANGSGPDQ 310

RESULT 5
PCT-US03-19834-2
; Sequence 2, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; TITLE OF INVENTION: METHODS FOR TREATING HYPERLIPIDEMIA
; FILE REFERENCE: IOWA-045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-2

Query Match      100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TDFDQGWGPISYANGSGPDQ 20
      |||||
Db      463 TDFDQGWGPISYANGSGPDQ 482

RESULT 6
US-10-296-734-406
; Sequence 406, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Hepc 1a consensus polyprotein
US-10-296-734-406

Query Match      100.0%; Score 20; DB 6; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TDFDQGWGPISYANGSGPDQ 20
      |||||
Db      463 TDFDQGWGPISYANGSGPDQ 482
```

```

; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 810
; LENGTH: 5985
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a savine
US-10-296-734-810

Query Match          95.0%; Score 19; DB 6; Length 5985;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 DFDQGWGPISYANGSGPDQ 20
      |||||||
Db      3961 DFDQGWGPISYANGSGPDQ 3979
      |||||||

RESULT 10
US-10-296-734-468
; Sequence 468, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296.734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 468
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 31
US-10-296-734-468

Query Match          80.0%; Score 16; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.1e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TDFDQGWGPISYANGS 16
      0 |||||
Db      15 TDFDQGWGPISYANGS 30
      |||||

RESULT 11
US-10-296-734-816
; Sequence 816, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296.734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 816
; LENGTH: 1997
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette C

US-10-296-734-470
; Sequence 470, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296.734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 470
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 32
US-10-296-734-470

Query Match          95.0%; Score 19; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 DFDQGWGPISYANGSGPDQ 20
      |||||||
Db      1 DFDQGWGPISYANGSGPDQ 19
      |||||||

RESULT 8
US-10-296-734-814
; Sequence 814, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296.734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 814
; LENGTH: 2010
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette B
US-10-296-734-814

Query Match          95.0%; Score 19; DB 6; Length 2010;
Best Local Similarity 100.0%; Pred. No. 6.9e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 DFDQGWGPISYANGSGPDQ 20
      |||||||
Db      1951 DFDQGWGPISYANGSGPDQ 1969
      |||||||

RESULT 9
US-10-296-734-810
; Sequence 810, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296.734
```


Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 ANGSGPDQ 20

|||||

Db 1 ANGSGPDQ 8

Search completed: November 21, 2003, 22:12:54
Job time : 8.55 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:46:27 ; Search time 9.25 Seconds
(without alignments)
207.932 Million cell updates/sec

Title: US-09-973-025-77

Perfect score: 20

Sequence: 1 TDFDQGWGPISYANGSGPDQ 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_76:*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	513	A44150	structural protein
2	20	100.0	1	GNWVC3	genome polyprotein
3	14	70.0	716	JQ1366	polyprotein - hepa
4	13	65.0	3011	1 GNWVCH	genome polyprotein
5	11	55.0	3011	1 S40770	genome polyprotein
6	10	50.0	138	2 S24075	envelope protein -
7	10	50.0	138	2 S24074	envelope protein -
8	10	50.0	138	2 S24081	envelope protein -
9	9	45.0	138	2 S24080	envelope protein -
10	9	45.0	640	2 JQ1584	genome polyprotein
11	8	40.0	127	2 B48776	polyprotein (E2/NS
12	8	40.0	138	2 S24097	envelope protein -
13	8	40.0	138	2 S24096	envelope protein -
14	8	40.0	138	2 S24078	envelope protein -
15	8	40.0	138	2 S24073	envelope protein -
16	8	40.0	138	2 S24105	envelope protein -
17	8	40.0	138	2 S24094	envelope protein -
18	8	40.0	138	2 S24068	envelope protein -
19	8	40.0	138	2 S24069	envelope glycoprot
20	8	40.0	315	2 PS0164	envelope glycoprot
21	8	40.0	550	2 JH0711	genome polyprotein
22	8	40.0	782	2 S19876	genome polyprotein
23	8	40.0	782	2 S18031	genome polyprotein
24	8	40.0	782	2 S19875	genome polyprotein
25	8	40.0	3010	1 GNWVTC	genome polyprotein
26	8	40.0	3010	1 GNWVTW	genome polyprotein
27	7	35.0	504	2 T50983	probable pleiotrop
28	7	35.0	504	2 JQ1925	polyprotein - hepa
29	7	35.0	523	2 JQ1926	polyprotein - hepa

30	7	35.0	676	2 D96630	hypothetical prote
31	6	30.0	107	2 T42275	hypothetical prote
32	6	30.0	134	2 S24089	envelope protein -
33	6	30.0	135	2 S24086	envelope protein -
34	6	30.0	135	2 A48776	polyprotein (E2/NS
35	6	30.0	135	2 C96026	probable transpos
36	6	30.0	135	2 H95878	probable ISM2011-
37	6	30.0	135	2 A95263	TRM2011-2a transp
38	6	30.0	135	2 B95386	TRM2011-2a transp
39	6	30.0	135	2 C95422	TRM2011-2a transp
40	6	30.0	136	2 S24090	envelope protein -
41	6	30.0	136	2 S24091	envelope protein -
42	6	30.0	136	2 E95401	TRM2011-2a transp
43	6	30.0	137	2 S24088	envelope protein -
44	6	30.0	138	2 S24100	envelope protein -
45	6	30.0	138	2 S24084	envelope protein -

ALIGNMENTS

RESULT 1

A44150

structural protein - hepatitis C virus

C:Species: hepatitis C virus

C:Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 17-Nov-2000

C:Accession: A44150

R:Ching, W.M.; Wychowski, C.; Beach, M.J.; Wang, H.; Davies, C.L.; Carl, M.; Bradley, D.

Proc. Natl. Acad. Sci. U.S.A. 89, 3190-3194, 1992

A:Title: Interaction of immune sera with synthetic peptides corresponding to the structu

A:Reference number: A44150; MUID:92228749; PMID:1373489

A:Accession: A44150

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: genomic RNA

A:Residues: 1-513 <CHI>

C:Superfamily: hepatitis C virus genome polyprotein

Query Match 100.0%; Score 20; DB 2; Length 513;

Best Local Similarity 100.0%; Pred. No. 1.6e-14; Mismatches 0; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20

|||||

Db 463 TDFDQGWGPISYANGSGPDQ 482

RESULT 2

GNWVC3

genome polyprotein - hepatitis C virus (strain HCV-1)

N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001

C:Accession: A39166; PQ0403; PQ0404

R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi

Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991

A:Title: Genetic organization and diversity of the hepatitis C virus.

A:Reference number: A39166; MUID:91172826; PMID:1848704

A:Accession: A39166

A:Molecule type: mRNA

A:Residues: 1-3011 <CHO>

A:Cross-references: GB:M62321; NID:G329873; PIDN:AAA45676.1; PID:G329874

R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L

J. Gen. Virol. 73, 1131-1141, 1992

A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e

A:Reference number: PQ0393; MUID:92268871; PMID:1316939

A:Accession: PQ0403

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CHA>

A:Cross-references: DBJ:D10128

A:Experimental source: isolates E-b16

A:Accession: PQ0404

A>Status: preliminary

Db 463 TDFDQGWGPIS 473
|||||

RESULT 6

S24075

envelope protein - hepatitis C virus (fragment)
N:Contains: envelope protein gp35; envelope protein gp70
C:Species: hepatitis C virus
C:Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C:Accession: S24075
R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraisho, K.; Ohkoshi, J. Gen. Virol. 73, 1521-1525, 1992
A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A:Reference number: S24068; MUID:92230394; PMID:1314471
A:Accession: S24075
A:Molecule type: genomic RNA
A:Residues: 1-138 <KAT>
A:Cross-references: EMBL:X60564
A:Experimental source: isolate RE37B
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; glycoprotein; polyprotein
F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted
F:49,55,62,80,108/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 50.0%; Score 10; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.00048;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 FDOQGWGPIS 12

Db 97 FDOQGWGPIS 106
|||||

RESULT 7

S24074

envelope protein - hepatitis C virus (fragment)
N:Contains: envelope protein gp35; envelope protein gp70
C:Species: hepatitis C virus
C:Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C:Accession: S24074
R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraisho, K.; Ohkoshi, J. Gen. Virol. 73, 1521-1525, 1992
A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A:Reference number: S24068; MUID:92230394; PMID:1314471
A:Accession: S24074
A:Molecule type: genomic RNA
A:Residues: 1-138 <KAT>
A:Cross-references: EMBL:X60563
A:Experimental source: isolate RE37A
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; glycoprotein; polyprotein
F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted
F:49,55,62,80,108/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 50.0%; Score 10; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.00048;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 FDOQGWGPIS 12

Db 97 FDOQGWGPIS 106
|||||

RESULT 8

S24081

envelope protein - hepatitis C virus (fragment)

N:Contains: envelope protein gp35; envelope protein gp70
C:Species: hepatitis C virus
C:Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C:Accession: S24081
R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraisho, K.; Ohkoshi, J. Gen. Virol. 73, 1521-1525, 1992
A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A:Reference number: S24068; MUID:92230394; PMID:1314471
A:Accession: S24081
A:Molecule type: genomic RNA
A:Residues: 1-138 <KAT>
A:Cross-references: EMBL:X60569
A:Experimental source: isolate RE43
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; glycoprotein; polyprotein
F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted
F:49,55,62,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 50.0%; Score 10; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.00048;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPIS 10

Db 95 TDFDQGWGPIS 104
|||||

RESULT 9

S24080

envelope protein - hepatitis C virus (fragment)
N:Contains: envelope protein gp35; envelope protein gp70
C:Species: hepatitis C virus
C:Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C:Accession: S24080
R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraisho, K.; Ohkoshi, J. Gen. Virol. 73, 1521-1525, 1992
A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A:Reference number: S24068; MUID:92230394; PMID:1314471
A:Accession: S24080
A:Molecule type: genomic RNA
A:Residues: 1-138 <KAT>
A:Cross-references: EMBL:X60568
A:Experimental source: isolate RE42
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; glycoprotein; polyprotein
F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted
F:49,55,62,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 45.0%; Score 9; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.006;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 QGQGPISYA 13

Db 99 QGQGPISYA 107
|||||

RESULT 10

JQ1584

Genome polyprotein - hepatitis C virus (strain U.K.) (fragment)
N:Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural protein 3; envelope protein E3
C:Species: hepatitis C virus
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
C:Accession: JQ1584
R:Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J. J. Gen. Virol. 73, 1521-1525, 1992
A:Title: Cloning and sequencing of the structural region and expression of putative core

C;Keywords: envelope protein; glycoprotein; polypeptide

F;1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F;16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F;16-42/Region: hypervariable 1 #status predicted
F;106-112/Region: hypervariable 2 #status predicted
F;49,55,62,80/Binding site: carbohydrate (Asn) #status predicted

Query Match 40.0%; Score 8; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.076;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 FDOGWGPI 10
| | | | | | | |
DB 97 FDOGWGPI 104

RESULT 13

S24096

envelope protein - hepatitis C virus (fragment)
N;Contains: envelope protein gp35; envelope protein gp70
C;Species: hepatitis C virus
C;Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C;Accession: S24096; S78233
R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraio, K.; Ohkoshi, H.;
Virus Res. 22, 107-123, 1992
A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A;Reference number: S24068; MUID:92230394; PMID:1314471
A;Accession: S24096
A;Molecule type: genomic RNA
A;Residues: 1-138 <KAT>
A;Cross-references: EMBL:X60581
A;Experimental source: isolate RE58
R;Kato, N.
submitted to the EMBL Data Library, August 1991
A;Reference number: S78226
A;Accession: G578233
A;Molecule type: genomic RNA
A;Residues: 1-23,'A',25-40,'Q',42-67,'G',69-138 <KAW>
A;Cross-references: EMBL:X60581
C;Superfamily: hepatitis C virus genome polypeptide
C;Keywords: envelope protein; glycoprotein; polypeptide
F;1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F;16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F;16-42/Region: hypervariable 1 #status predicted
F;106-112/Region: hypervariable 2 #status predicted
F;49,55,62,80/Binding site: carbohydrate (Asn) #status predicted

Query Match 40.0%; Score 8; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.076;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 FDOGWGPI 10
| | | | | | | |
DB 97 FDOGWGPI 104

RESULT 14

S24078

envelope protein - hepatitis C virus (fragment)
N;Contains: envelope protein gp35; envelope protein gp70
C;Species: hepatitis C virus
C;Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C;Accession: S24078
R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraio, K.; Ohkoshi, H.;
Virus Res. 22, 107-123, 1992
A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A;Reference number: S24068; MUID:92230394; PMID:1314471
A;Accession: S24078
A;Molecule type: genomic RNA
A;Residues: 1-138 <KAT>
A;Cross-references: EMBL:X60566
A;Experimental source: isolate RE40
C;Superfamily: hepatitis C virus genome polypeptide

C:Keywords: envelope protein; glycoprotein; polyprotein
F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted
F:49,55,62,80,110/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 40.0%; Score 8; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.076;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 FDQGWGPI 10
|||
Db 97 FDQGWGPI 104

RESULT 15
S24073
envelope protein - hepatitis C virus (fragment)
N:Contains: envelope protein gp35; envelope protein gp70
C:Species: hepatitis C virus
C:Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C:Accession: S24073
R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; MuraIso, K.; Ohkoshi,
Virus Res. 22, 107-123, 1992
A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A:Reference number: S24068; PMID:92230394; PMID:1314471
A:Accession: S24073
A:Molecule type: genomic RNA
A:Residues: 1-138 <MAT>
A:Cross-references: EMBL:X60562
A:Experimental source: isolate RE36
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; glycoprotein; polyprotein
F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted
F:49,55,62,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 40.0%; Score 8; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.076;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 FDQGWGPI 10
|||
Db 97 FDQGWGPI 104

Search completed: November 21, 2003, 21:11:31
Job time : 10.25 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:52 ; Search time 4.85 seconds
(without alignments)
193.925 Million cell updates/sec

Title: US-09-973-025-77

Perfect score: 20

Sequence: 1 TDFDQMGPISYANGSPDQ 20

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	20	100.0	1 POLG_HCV1	P26664 h genome po
2	13	65.0	1 POLG_HCVH	P27958 h genome po
3	8	40.0	1 POLG_HCVJ2	P27959 hepatitis c
4	8	40.0	1 POLG_HCVBK	P26663 h genome po
5	8	40.0	1 POLG_HCVTW	P29846 h genome po
6	7	35.0	1 POLG_HCVH4	Q01404 hepatitis c
7	7	35.0	1 POLG_HCVHK	Q01403 hepatitis c
8	6	30.0	1 TL29_LYCES	Q9thx6 lycopersico
9	6	30.0	1 TL29_ARATH	P82281 arabidopsis
10	6	30.0	1 CTPT_PLAFK	P49587 plasmodium
11	6	30.0	1 EVX2_HETFR	Q91a18 heterodontu
12	6	30.0	1 STCL_EMENI	Q00707 emericeila
13	6	30.0	1 APH4_DROME	Q24238 drosophila
14	6	30.0	1 CIX5_MUSPF	P79197 mustela put
15	6	30.0	1 DD17_HUMAN	Q92841 homo sapien
16	6	30.0	1 PANG_DROME	P91943 drosophila
17	6	30.0	1 KNBR_YEAST	P53739 saccharomyc
18	6	30.0	1 WR52_ARATH	Q9fh83 arabidopsis
19	6	30.0	1 POLG_HCVJA	P26662 h genome po
20	5	25.0	1 POLG_HCVJT	Q00269 h genome po
21	5	25.0	1 NUIJ_COTJA	P24968 coturnix co
22	5	25.0	1 RELX_DASSA	P81191 dasvatis sa
23	5	25.0	1 ELIA_PHYCP	P15571 phytophthor
24	5	25.0	1 ELIA_PHYME	P35698 phytophthor
25	5	25.0	1 GLPE_YERPE	Q8xj13 yersinia pe
26	5	25.0	1 EL11_PHYCR	P41802 phytophthor
27	5	25.0	1 EL1B_PHYCR	P15570 phytophthor
28	5	25.0	1 PSN_HUMAN	O60542 homo sapien
29	5	25.0	1 ISPF_YERPE	Q82bp7 yersinia pe
30	5	25.0	1 ATPF_ENTHR	P26681 enterococcu
31	5	25.0	1 NUSG_LACLA	Q9cgv7 lactococcus
32	5	25.0	1 RACC_DICDI	P34149 dictyosteli
33	5	25.0	1 VE26_NPVOP	O10276 orgyia pseu

34 5 25.0 199 1 SOXE_SULAC Q53765 sulfolobus
35 5 25.0 199 1 YG4T_YEAST P50083 saccharomyc
36 5 25.0 200 1 R102_YEAST O2721 saccharomyc
37 5 25.0 202 1 UMPK_ARATH O04905 arabidopsis
38 5 25.0 206 1 SOD2_PLEBO P50059 plectonema
39 5 25.0 211 1 WIN2_SOLFU P09762 solanum tub
40 5 25.0 212 1 KTHY_BACSU P37537 bacillus su
41 5 25.0 217 1 OS25_PLARE P19455 plasmodium
42 5 25.0 222 1 OUL3_MOUSE P34984 mus musculu
43 5 25.0 222 1 YAGY_ECOLI P77188 escherichia
44 5 25.0 225 1 SRD1_YEAST P09007 saccharomyc
45 5 25.0 231 1 RNH_STRCO Q9x7r6 streptomyce

ALIGNMENTS

RESULT 1
POLG_HCV1 0 STANDARD; PRT; 3011 AA.
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein c (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate 1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=911172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA Bradley D.W., Kuo G., Houghton M.;
RT "Genetic organization and diversity of the hepatitis C virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M62321; AAA45676.1; -.
DR PIR; A39166; GNMVC3.
DR PDB; 1A1V; 16-FEB-99.
DR PDB; 1HEI; 25-NOV-98.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.

DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA pol DS PS.
DR InterPro; IPR007094; RNA pol Psvir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01539; HCV core; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; helicase C; 1.
DR Pfam; PF00998; Viral RdRp; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 283 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;
Query Match 100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 7.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFDQGWGPISYANGSGPDQ 20

Db 463 TDFDQGWGPISYANGSGPDQ 482
RESULT 2
POLG HCVH STANDARD; PRT; 3011 AA.
AC P27958; 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate H) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658800;
RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C
RT virus: comparison with American and Japanese isolates.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX MEDLINE=97331322; PubMed=9187654;
RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain.";
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE=98154321; PubMed=9493270;
RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RL of unwinding.";
Structure 6:99-100(1998).
CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC [RNA] (N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.

CC DR EMBL; M67463; AAA45534.1; -
DR PIR; A36814; GNMVCH.
DR PDB; 1HEI; 25-NOV-98.
DR PDB; 1A1V; 16-FEB-99.
DR PDB; 1A1R; 17-JUN-98.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR TRANSFAC; T04155; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RGRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA pol DS ps.
DR InterPro; IPR007094; RNA pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase C; 1.
DR Pfam; PF00998; Viral RdRp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
CELLULAR AMINOPEPTIDASE.
FT CHAIN 1 191 CAPSID PROTEIN C.
FT CHAIN 192 383 ENVELOPE GLYCOPROTEIN E1.
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E2.
FT CHAIN 747 809 PROTEIN P7.
FT CHAIN 810 1026 NONSTRUCTURAL PROTEIN NS2.
FT CHAIN 1027 1657 PROTEASE/HELICASE NS3.
FT CHAIN 1658 1711 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS5A.
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5B.
FT CHAIN 3011 369 POTENTIAL.
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT STRAND 1224 1226
FT TURN 1232 1233
FT TURN 1236 1238
FT HELIX 1239 1246
FT TURN 1247 1248
FT STRAND 1251 1255
FT HELIX 1258 1271
FT TURN 1272 1272
FT STRAND 1277 1280
FT TURN 1281 1282
FT STRAND 1283 1285
FT STRAND 1291 1295
FT HELIX 1296 1301
FT TURN 1302 1303
FT STRAND 1312 1316
FT TURN 1317 1319
FT HELIX 1323 1335
FT TURN 1336 1340
FT STRAND 1343 1347
FT TURN 1352 1353
FT STRAND 1360 1361
FT STRAND 1362 1366
FT STRAND 1368 1368
FT STRAND 1373 1375
FT TURN 1376 1377
FT STRAND 1378 1380
FT HELIX 1382 1385
FT STRAND 1389 1393
FT HELIX 1397 1409
FT TURN 1410 1411
FT STRAND 1414 1417
FT TURN 1419 1420
FT STRAND 1432 1436
FT TURN 1438 1439
FT STRAND 1450 1453
FT STRAND 1456 1463
FT STRAND 1471 1478
FT STRAND 1480 1480
FT HELIX 1481 1488
FT TURN 1489 1490
FT STRAND 1497 1501
FT STRAND 1507 1507
FT STRAND 1511 1511
FT HELIX 1514 1527
FT HELIX 1532 1544
FT STRAND 1550 1550
FT HELIX 1555 1564
FT HELIX 1570 1578
FT TURN 1579 1580
FT HELIX 1584 1597
FT TURN 1598 1598
FT HELIX 1606 1611
FT TURN 1614 1618
FT STRAND 1622 1623
FT STRAND 1627 1627
FT STRAND 1635 1636
FT HELIX 1640 1652
SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;
Query Match 65.0%; Score 13; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred.No. 2,7e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 5 QGWGPISYANGSG 17
| | | | | | | | | | | | | | | | | | | | | |
Db 467 QGWGPISYANGSG 479
RESULT 3
POLG_HCVJ2
ID POLG_HCVJ2 STANDARD; PRT; 513 AA.
AC P27959;
DT 01-AUG-1992 (Rel. 23, Created)

01-AUG-1992 (Rel. 23, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 (GP68) (GP70) (NS1)] (Fragment).
 Hepatitis C virus (isolate HC-J2) (HCV).
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.
 NCBI_TaxID=11111;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=92230232; PubMed=1314459;
 Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 Tanaka T., Fukuda S., Tada F., Mishiro S.;
 "Full-length sequence of a hepatitis C virus genome having poor
 homology to reported isolates: comparative study of four distinct
 genotypes.";
 J. Virol. 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D10074; BAA00968.1; -;
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR PolyProtein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 >513
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT NON_TER 513 513
 SQ SEQUENCE 513 AA; 55704 MW; 943F31E3514CDEF3 CRC64;
 Query Match 40.0%; Score 8; DB 1; Length 513;
 Best Local Similarity 100.0%; Pred. No. 0.14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 3 FDQGWGPI 10
 |||||
 Db 465 FDQGWGPI 472

RESULT 4
 POLG HCVBK 0 STANDARD; PRT; 3010 AA.
 ID POLG HCVBK
 AC P26663;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate BK) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11105;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=91140698; PubMed=1847440;
 RP Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 RP Onishi E., Andoh T., Yoshida I., Okayama H.;
 RA "Structure and organization of the hepatitis C virus genome isolated
 RT from human carriers.";
 RL J. Virol. 65:1105-1113(1991).
 RN [2]
 RP SEQUENCE OF 1487-1500.
 RP MEDLINE=96235224; PubMed=8647104;
 RP Borowski P., Heiland M., Oehlmann K., Becker B., Kornetky L.;
 RA "Non-structural protein 3 of hepatitis C virus inhibits
 RT phosphorylation mediated by cAMP-dependent protein kinase.";
 RL Eur. J. Biochem. 237:611-618(1996).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RP MEDLINE=97015088; PubMed=8861916;
 RA Love R.A., Farge H.E., Wickham J.A., Hostomsky Z., Habuka N.,
 RA Moomaw E.W., Adachi T., Hostomska Z.;
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 RL trypsin-like fold and a structural zinc binding site.";
 RN Cell 87:331-342(1996).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RP MEDLINE=98227846; PubMed=9568891;
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
 RL virus: a 2-2-A resolution structure in a hexagonal crystal form.";
 CC Protein Sci. 7:837-847(1998)
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

DR EMBL; M58335; AAA72945.1; --
 DR PIR; A38465; GNMVTC.
 DR PDB; 1A1Q; 25-MAR-98.
 DR PDB; 1JXP; 14-JAN-98.
 DR PDB; 1NS3; 08-APR-98.
 DR PDB; 1C2P; 15-NOV-00.
 DR PDB; 1CSJ; 08-NOV-99.
 DR PDB; 1GX5; 09-APR-02.
 DR PDB; 1GX6; 10-APR-02.
 DR PDB; 1QUV; 26-JUN-00.
 DR PDB; 8OHM; 20-APR-99.
 DR MEROPS; S29.001; --
 DR MEROPS; U39.001; --
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NSI.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RGRP.
 DR InterPro; IPR007095; RNA_pol_DS_Ps.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NSI; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00998; Viral_RGRP; 1.
 DR ProDom; PD186062; HCV_NSI; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1107
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1319 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623

FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2529 2529
 FT CARBOHYD 2788 2788
 FT STRAND 1031 1035
 FT HELIX 1039 1047
 FT STRAND 1050 1050
 FT STRAND 1059 1063
 FT STRAND 1068 1074
 FT TURN 1075 1076
 FT STRAND 1077 1081
 FT HELIX 1082 1085
 FT TURN 1086 1087
 FT STRAND 1090 1092
 FT TURN 1093 1094
 FT STRAND 1095 1097
 FT STRAND 1101 1103
 FT TURN 1104 1107
 FT STRAND 1108 1112
 FT STRAND 1120 1120
 FT STRAND 1122 1122
 FT STRAND 1129 1133
 FT TURN 1135 1136
 FT STRAND 1139 1144
 FT STRAND 1149 1157
 FT HELIX 1158 1161
 FT TURN 1162 1163
 FT STRAND 1165 1166
 FT TURN 1168 1171
 FT TURN 1172 1174
 FT STRAND 1175 1186
 FT TURN 1187 1188
 FT STRAND 1189 1197
 FT HELIX 1198 1202
 FT TURN 1203 1204
 FT STRAND 1680 1688
 SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCDFD9C CRC64;

N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 40.0%; Score 8; DB 1; Length 3010;

Best Local Similarity 100.0%; Pred. No. 0.7;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 FDCQWGPI 10

Db 465 FDCQWGPI 472

RESULT 5

POLG HCVTW STANDARD; PRT; 3010 AA.

AC P29846;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);

DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)

DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)

DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein

DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48);

OS Hepatitis C virus (isolate Taiwan) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=31645;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92230206; PubMed=1314449;

RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;

RT "The Taiwanese hepatitis C virus genome: sequence determination and

mapping the 5' termini of viral genomic and antigenomic RNA.";

Virology 188:102-113(1992).

-|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.

-|- NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

-|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polypeptide, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.

-|- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA) (N).

-|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.

-|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M84754; -; NOT ANNOTATED_CDS.

DR PIR; A40244; GNMVTV.

DR PDB; 1N64; 25-FEB-03.

DR PDB; 1NS3; 08-APR-98.

DR MEROPS; S29.001; -.

DR MEROPS; U39.001; -.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV capsid.

DR InterPro; IPR002521; HCV core.

DR InterPro; IPR002519; HCV env.

DR InterPro; IPR002531; HCV NS1.

DR InterPro; IPR002518; HCV NS2.

DR InterPro; IPR004109; HCV NS3.

DR InterPro; IPR00745; HCV NS4a.

DR InterPro; IPR001490; HCV NS4b.

DR InterPro; IPR002868; HCV NS5a.

DR InterPro; IPR002166; HCV NS5b.

DR InterPro; IPR007095; RNA pol ps.

DR InterPro; IPR007094; RNA pol psvir.

DR Pfam; PF01543; HCV capsid; 1.

DR Pfam; PF01542; HCV core; 1.

DR Pfam; PF01539; HCV env; 1.

DR Pfam; PF01560; HCV NS1; 1.

DR Pfam; PF01538; HCV NS2; 1.

DR Pfam; PF02907; HCV NS3; 1.

DR Pfam; PF01006; HCV NS4a; 1.

DR Pfam; PF01001; HCV NS4b; 1.

DR Pfam; PF01506; HCV NS5a; 1.

DR Pfam; PF00271; Helicase_C; 1.

DR Pfam; PF00998; Viral RdRP; 1.

DR ProDom; PD186062; HCV NS1; 1.

DR SMART; SM00487; DEXDC; 1.

KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;

KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;

KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;

KW 3D-structure.

INIT_MET 1 1

REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE.

FT CHAIN 1 115

FT CHAIN 116 191

FT CHAIN 192 383

FT CHAIN 384 729

FT CHAIN 730 1006

FT CHAIN 1007 1615

FT CHAIN 1616 1862

FT CHAIN 1863 2013

FT CHAIN 2014 3010

FT CHAIN 3010 369

FT TRANSMEM 347 1083

FT ACT_SITE 1083 1107

FT ACT_SITE 1107 1107

CHARGE RELAY SYSTEM (BY SIMILARITY).

ACT SITE 1165 1165

NP BIND 1230 1237

SITE 1316 1319

CARBOHYD 196 196

CARBOHYD 209 209

CARBOHYD 233 233

CARBOHYD 234 234

CARBOHYD 250 250

CARBOHYD 305 305

CARBOHYD 417 417

CARBOHYD 423 423

CARBOHYD 430 430

CARBOHYD 448 448

CARBOHYD 532 532

CARBOHYD 540 540

CARBOHYD 556 556

CARBOHYD 576 576

CARBOHYD 623 623

CARBOHYD 645 645

CARBOHYD 2041 2041

CARBOHYD 2077 2077

CARBOHYD 2240 2240

CARBOHYD 2529 2529

CARBOHYD 2788 2788

SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 40.08; Score 8; DB 1; Length 3010;

Best Local Similarity 100.0%; Pred. No. 0.7;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 FDQGWGPI 10

Db 465 FDQGWGPI 472

RESULT 6

POLG HCVH4

ID POLG HCVH4 STANDARD; PRT; 520 AA.

AC Q01404; 0

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1)] (Fragment).

DE Hepatitis C virus (isolate HCV-476) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=31643;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93019030; PubMed=1383400;

RA Abe K., Inchauspe G., Fujisawa K.;

RT "Genomic characterization and mutation rate of hepatitis C virus isolated from a patient who contracted hepatitis during an epidemic of non-A, non-B hepatitis in Japan.";

RL J. Gen. Virol. 73:2725-2729(1992).

CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; D10688; BAA01530.1; -.

DR InterPro; IPR002522; HCV capsid.

DR InterPro; IPR002521; HCV core.

DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 >520
 FT TRANSMEM 347 369
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 418 418
 FT CARBOHYD 424 424
 FT CARBOHYD 431 431
 FT CARBOHYD 449 449
 FT NON_TER 520 520
 SQ SEQUENCE 520 AA; 56499 MW; AA135246CF20D525 CRC64;
 Query Match 35.0%; Score 7; DB 1; Length 520;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 3 FDOQWGP 9
 Db 466 FDOQWGP 472

RESULT 7
 POLG HCVHK
 ID POLG HCVHK STANDARD; PRT; 520 AA.
 AC Q01403;
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1) (Fragment).
 OS Hepatitis C virus (isolate HCV-KF) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 CC NCBI_TaxID=31644;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93019030; PubMed=1383400;
 RA Abe K., Inchauspe G., Fujisawa K.;
 RT "Genomic characterization and mutation rate of hepatitis C virus
 RT isolated from a patient who contracted hepatitis during an epidemic
 RT of non-A, non-B hepatitis in Japan."
 RL J. Gen. Virol. 73:2725-2729(1992).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; D10687; BAA01529.1; -;
 DR PIR; JQ1925; JQ1925.

DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT INIT_MET 1 1
 FT CHAIN 0 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 >520
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 418 418
 FT CARBOHYD 424 424
 FT CARBOHYD 431 431
 FT CARBOHYD 449 449
 FT NON_TER 520 520
 SQ SEQUENCE 520 AA; 56476 MW; 1D2BD0A6FF27349B CRC64;
 Query Match 35.0%; Score 7; DB 1; Length 520;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 3 FDOQWGP 9
 Db 466 FDOQWGP 472

RESULT 8
 TL29 LYCES
 ID TL29 LYCES STANDARD; PRT; 345 AA.
 AC Q9THX6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative L-ascorbate peroxidase, chloroplast precursor (EC 1.11.1.11)
 DE (thylakoid lumenal 29 kDa protein) (TL29) (P29).
 GN CLEB3J9.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asterididae; lamids; Solanales; Solanaceae; Solanum.
 CC NCBI_TaxID=4081;
 RN [1]
 RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
 RX MEDLINE=20487156; PubMed=11034343;
 RA Kieselbach T., Bystedt M., Hynds P., Robinson C., Schroeder W.P.;
 RT "A peroxidase homologue and novel plastocyanin located by proteomics
 RT to the Arabidopsis chloroplast thylakoid lumen."
 RL FEBS Lett. 480:271-276(2000).
 CC -!- FUNCTION: PLAYS A KEY ROLE IN HYDROGEN PEROXIDE REMOVAL IN
 CC THE CHLOROPLASTS AND CYTOSOL OF HIGHER PLANTS.
 CC -!- CATALYTIC ACTIVITY: L-ascorbate + H(2)O(2) = dehydroascorbate + 2
 CC H(2)O.
 CC -!- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.
 CC -!- SIMILARITY: TO ASCORBATE PEROXIDASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; D10687; BAA01529.1; -;
 DR PIR; JQ1925; JQ1925.

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC EMBL; AJ251892; CAB64343.1; --
CC HSSP; P48534; IAPX.
CC InterPro; IPR002016; Peroxidase.
CC Pfam; PF00141; peroxidase; 1.
KW Oxidoreductase; Peroxidase; Hydrogen peroxide; Chloroplast;
KW Transit peptide; Thylakoid.
FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
FT TRANSIT 78 THYLAKOID (BY SIMILARITY).
FT CHAIN 79 345 PUTATIVE L-ASCORBATE PEROXIDASE.
SQ SEQUENCE 345 AA; 37749 MW; 681CABBFDDBB841B CRC64;

Query Match 30.0%; Score 6; DB 1; Length 345;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GPSIYA 13
DB 171 GPSIYA 176

RESULT 9
TL29 ARATH
ID TL29 ARATH STANDARD; PRT; 349 AA.
AC P82281; Q9M0S6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Putative L-ascorbate peroxidase, chloroplast precursor (EC 1.11.1.11)
DE (Thylakoid lumenal 29 kDa protein) (TL29) (P29).
GN AT4G09010 OR F23J3.40.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsidae.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Columbia;
RC MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzenecker T., Bothe G., Ransperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren N., Dirkse W.,
RA Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Berneriser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buysaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allein S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McClay K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloeker H., Schaefer M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maare A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Partmann B., Graeber K., Dauner D., Herzl A.,
RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chedof F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bieleke C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,

```

```

RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spieth J., Ryan K., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E.P., Marra M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:769-777(1999).
RN [2]
RP SEQUENCE OF 83-112, AND SUBCELLULAR LOCATION.
RC STRAIN=cv. Columbia;
RX MEDLINE=20487156; PubMed=11034343;
RA Kieselbach T., Bystedt M., Hynds P., Robinson C., Schroeder W.P.;
RT "A peroxidase homologue and novel plastocyanin located by proteomics
RT to the Arabidopsis chloroplast thylakoid lumen.";
RL FEBS Lett. 480:271-276(2000).
CC -!- FUNCTION: PLAYS A KEY ROLE IN HYDROGEN PEROXIDE REMOVAL IN
CC THE CHLOROPLASTS AND CYTOSOL OF HIGHER PLANTS.
CC -!- CATALYTIC ACTIVITY: L-ascorbate + H(2)O(2) = dehydroascorbate + 2
CC H(2)O.
CC -!- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.
CC -!- SIMILARITY: TO ASCORBATE PEROXIDASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

CC EMBL; AL161513; CAB78025.1; --
DR PIR; A85091; A85091.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
DR PROSITE; PS00873; PEROXIDASE 4; 1.
KW Oxidoreductase; Peroxidase; Hydrogen peroxide; Chloroplast;
KW Transit peptide; Thylakoid.
FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
FT TRANSIT 82 THYLAKOID.
FT CHAIN 83 349 PUTATIVE L-ASCORBATE PEROXIDASE.
SQ SEQUENCE 349 AA; 37934 MW; ABD439188BDA8E61 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 349;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GPSIYA 13
DB 175 GPSIYA 180

RESULT 10
CTPT PLAPK
ID CTPT PLAPK STANDARD; PRT; 370 AA.
AC P49587;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cholinephosphate cytidylyltransferase (EC 2.7.7.15) (Phosphorylcholine
DE transferase) (CTP:phosphocholine cytidylyltransferase) (CT) (CTP).
GN CTP.
OS Plasmodium falciparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5839;

```

RN SEQUENCE FROM N.A.
 RP MEDLINE=96061933; PubMed=7588775;
 RX Yeo H.-J., Sri Widada J., Mercereau-Puijalon O., Vial H.J.;
 RT "Molecular cloning of CTP:phosphocholine cytidylyltransferase from
 RT Plasmodium falciparum.";
 RL Eur. J. Biochem. 233:62-72(1995).
 CC -!- FUNCTION: CONTROLS PHOSPHATIDYLCHOLINE SYNTHESIS.
 CC -!- CATALYTIC ACTIVITY: CTP + choline phosphate = diphosphate + CDP-
 CC choline.
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE ASEQUAL INTRAERYTHROCYTIC
 CC STAGES.
 CC -!- SIMILARITY: Belongs to the cytidylyltransferase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X84041; CAA58860.1; -.
 DR PIR: S68187; S68187.
 DR HSP: P27623; ICOZ.
 DR InterPro: IPR004821; Cyt tran rel.
 DR InterPro: IPR004820; Cytidyltransf.
 DR Pfam: PF01467; CTP transf.2; 1.
 DR TIGRFAMs: TIGR00125; cyt_tran_rel; 1.
 KW Transferase; Nucleotidyltransferase; Phospholipid biosynthesis.
 FT DOMAIN 89 268 CATALYTIC (POTENTIAL).
 FT DOMAIN 57 64 POLY-ASP.
 FT DOMAIN 193 203 POLY-LYS.
 FT SEQUENCE 370 AA; 42630 MW; B2E457D09B06FB6 CRC64;
 SQ
 Query Match 30.0%; Score 6; DB 1; Length 370;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDFDQG 6
 DB 302 TDFDQG 307
 RESULT 11
 EVX2_HETFR STANDARD; PRT; 424 AA.
 AC Q9TA18;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Homeobox even-skipped homolog protein 2 (EVX-2).
 GN EVX2.
 OS Heterodontus francisci (Horn shark).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;
 OC Heterodontidae; Heterodontus.
 OX NCBI_TaxID=7792;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20144096; PubMed=10677514;
 RA Kim C.B., Amemiya C., Bailey W., Kawasaki K., Mezey J., Miller W.,
 RA Minoshima S., Shimizu N., Wagner G., Ruddle F.;
 RT "Hox cluster genomics in the horn shark, Heterodontus francisci.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:1655-1660(2000).
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE EVEN-SKIPPED HOMEBOX FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AF224263; AAP44638.1; -.
 DR HSP: P14653; IB72.
 DR TRANSFAC: T04488; -.
 DR InterPro: IPR001356; Homeobox.
 DR Pfam: PF00046; homeobox; 1.
 DR ProDom: PD00010; Homeobox; 1.
 DR SMART: SM00389; HOX; 1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS00071; HOMEBOX_2; 1.
 KW DNA-binding; Developmental protein; Homeobox; Nuclear protein.
 FT DNA BIND 179 238 HOMEBOX.
 FT DOMAIN 285 292 POLY-ALA.
 FT DOMAIN 340 353 POLY-ALA.
 SQ SEQUENCE 424 AA; 45629 MW; DA0DAF3E35D640E CRC64;
 Query Match 30.0%; Score 6; DB 1; Length 424;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 15 GSGPDQ 20
 DB 173 GSGPDQ 178
 RESULT 12
 STCL_EMENI STANDARD; PRT; 500 AA.
 ID STCL_EMENI
 AC Q00707;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable sterigmatocystin biosynthesis P450 monooxygenase STCL
 DE (EC 1.14.-.-) (Cytochrome P450 60B).
 GN STCL OR CYP60B.
 OS Emericella nidulans (Aspergillus nidulans).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; Emericella.
 OX NCBI_TaxID=162425;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FGSC 26;
 RX MEDLINE=96202293; PubMed=8643646;
 RA Brown D.W., Yu J.-H., Kellar H.S., Fernandes M., Nesbitt T.C.,
 RA Keller N.P., Adams T.H., Leonard T.J.;
 RT "Twenty-five coregulated transcripts define a sterigmatocystin gene
 RT cluster in Aspergillus nidulans.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422(1996).
 CC -!- FUNCTION: MAY BE INVOLVED IN THE HYDROXYLATION OF AVERANTIN TO
 CC FORM 5'-HYDROXYAVERANTIN, OXIDATION OF AVERUFIN TO 1-
 CC HYDROXYVERSICOLORONE, AND ITS OXIDATION TO VERSICONAL HEMIACETAL
 CC ACETATE.
 CC -!- PATHWAY: Sterigmatocystin biosynthesis.
 CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U34740; AAC49200.1; -.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Heme.
 FT METAL 444 444 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).

SQ SEQUENCE 500 AA; 56760 MW; BOF4D7E93DA3B445 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 500;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 6; Conservative 0; Mismatches 0; Gaps 0;

OY 10 ISYANG 15
 |||||
 DB 82 ISYANG 87

RESULT 13

AP44 DROME STANDARD; PRT; 596 AA.

AC Q24238; QIIM40; Q8SXW6; Q9VA19;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Alkaline phosphatase 4 precursor (EC 3.1.3.1).
 GN APH-4 OR CG1462.

OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;

RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM A), FUNCTION, TISSUE SPECIFICITY, AND
 RP DEVELOPMENTAL STAGE.

RC TISSUE=Head;
 RX MEDLINE=20094770; PubMed=1062988;
 RA Yang M.Y., Wang Z., MacPherson M., Dow J.A.T., Kaiser K.;
 FT "A novel Drosophila alkaline phosphatase specific to the ellipsoid
 RT body of the adult brain and the lower Malpighian (renal) tubule.";
 RL Genetics 154:285-297 (2000).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieri S., Fleischmann W.,
 RA Foaier C., Gabrielson A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milehina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinschenk G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RN Science 287:2185-2195 (2000).
 RN [3]

RP REVISIONS, AND ALTERNATIVE SPLICING.
 RC STRAIN=Berkley;
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochownik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review.";
 RT Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).
 RN [4]

RP SEQUENCE FROM N.A. (ISOFORM A).
 RC STRAIN=Berkley; TISSUE=Head;
 RX MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarin H., Kronmiller B., Pacle J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celniker S.E.;
 RT "A Drosophila full-length cDNA resource.";
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).

CC -!- FUNCTION: Important role in neural and renal epithelial function.
 CC -!- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an
 CC alcohol + phosphate.
 CC -!- SUBUNIT: Homodimer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
 CC similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=A;
 CC IsoId=Q24238-1; Sequence=Displayed;
 CC Name=B;
 CC IsoId=Q24238-2; Sequence=VSP_007002;
 CC Note=No experimental confirmation available;
 CC -!- TISSUE SPECIFICITY: Ellipsoid body ring neurons in the adult brain
 CC and in the lower Malpighian tubule and ureter.
 CC -!- DEVELOPMENTAL STAGE: Highest abundance during larval stage (prior
 CC to the secretion of pupal cuticle) and adult stage.
 CC -!- SIMILARITY: BELONGS TO THE ALKALINE PHOSPHATASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/submit/> or
 CC or send an email to license@ebi.ac.uk).

CC EMBL; X98402; CAA67052.1; -;
 CC EMBL; AE003776; AAF51106.1; -;
 CC EMBL; AE003776; AAN14265.1; -;
 CC EMBL; AY075544; AAL68351.1; -;
 CC FlyBase; FBgn0016123; Aph-4.
 CC InterPro; IPR001952; Alk_phosphatase.
 CC Pfam; PF00245; alk_phosphatase; 1.
 CC PRINTS; OPR00113; ALKPHTASE.
 CC ProDom; PD001868; Alk_phosphatase; 2.
 CC SMART; SM00098; alkpgc_1.
 CC PROSITE; PS00123; ALKALINE_PHOSPHATASE; 1.
 CC HydroLase; Zinc; Magnesium; Phosphorylation; Transmembrane;
 KW Glycoprotein; GPI-anchor; Signal; Alternative splicing.
 FT SIGNAL 1 20
 FT CHAIN 1 570 ALKALINE PHOSPHATASE 4.
 FT PROPEP 571 596 REMOVED IN MATURE FORM.
 FT DISULFID 539 550 BY SIMILARITY.
 FT LIPID 570 570 GPI-ANCHOR (POTENTIAL).
 FT TRANSMEM 571 591 POTENTIAL.
 FT MOD_RES 144 144 PHOSPHORYLATION (BY SIMILARITY).


```

FT ACT SITE 144 144 BY SIMILARITY.
FT METAL 92 92 MAGNESIUM (BY SIMILARITY).
FT METAL 93 93 ZINC 2 (BY SIMILARITY).
FT METAL 202 202 MAGNESIUM (BY SIMILARITY).
FT METAL 204 204 MAGNESIUM (BY SIMILARITY).
FT METAL 369 369 MAGNESIUM (BY SIMILARITY).
FT METAL 374 374 ZINC 1 (BY SIMILARITY).
FT METAL 378 378 ZINC 1 (BY SIMILARITY).
FT METAL 415 415 ZINC 2 (BY SIMILARITY).
FT METAL 416 416 ZINC 2 (BY SIMILARITY).
FT METAL 418 418 ZINC 1 (BY SIMILARITY).
FT METAL 504 504 ZINC 1 (BY SIMILARITY).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 401 401 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 464 464 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 1 94 Missing (in isoform B).
FT CONFLICT 200 200 /FTID=V8Pf 007002.
FT CONFLICT 358 358 I -> N (IN REF. 4).
FT CONFLICT 375 375 G -> D (IN REF. 1).
FT CONFLICT 495 495 G -> H (IN REF. 1).
FT CONFLICT 538 538 AT -> EP (IN REF. 1).
FT GCLGPAKDFDSCEDHKGQKDRPLDKPNKNGATVVGAS
FT LIPILTAATAILRGRL -> AVWVPRTSMTVPIRTRMG
FT KNIGRTWNPQIEVAPLLWEPP (IN REF. 1).
SQ SEQUENCE 596 AA; 65262 MW; 333F3345BFEAFB CRC64;

Query Match 30.0%; Score 6; DB 1; Length 596;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 ISYANG 15
Db 450 ISYANG 455

RESULT 14
C1KS MUSPF STANDARD; PRT; 601 AA.
ID C1KS MUSPF STANDARD; PRT; 601 AA.
AC P79197;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Potassium voltage-gated channel subfamily A member 5 (Potassium
DE channel Kv1.5).
GN KCNA5.
OS Muscota putorius furo (Ferret).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustelinae;
OC Mustela.
OX NCBI_TaxID=9669;
RN [1]_TaxID=9669;
RP SEQUENCE FROM N.A.
RC TISSUE=Heart atrium;
RA Schwegel T., Folander K., Swanson R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Mediates the voltage-dependent potassium ion
CC permeability of excitable membranes. Assuming opened or closed
CC conformations in response to the voltage difference across the
CC membrane, the protein forms a potassium-selective channel through
CC which K(+) ions may pass in accordance with their electrochemical
CC gradient (By similarity).
CC -!- SUBUNIT: Heterotrimer of potassium channel proteins (Probable).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- DOMAIN: The amino terminus may be important in determining the
CC rate of inactivation of the channel while the tail may play a role
CC in modulation of channel activity and/or targeting of the channel
CC to specific subcellular compartments.
CC -!- DOMAIN: The segment S4 is probably the voltage-sensor and is
CC characterized by a series of positively charged amino acids at
CC every third position.
CC -!- SIMILARITY: Belongs to the potassium channel family. A (Shaker)

```

```

CC subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U45979; AAB41145.1; -.
CC HSSP; Q54397; IBL8.
CC InterPro; IPR000210; BTB_POZ.
CC InterPro; IPR005821; Ion_trans.
CC InterPro; IPR001622; K-channel_pore.
CC InterPro; IPR003091; K_channel.
CC InterPro; IPR003131; K_tetra.
CC InterPro; IPR004052; KV15channel.
CC InterPro; IPR003968; Kv_channel.
CC InterPro; IPR005820; M-channel_nlg.
CC InterPro; IPR003972; Shaker_channel.
CC Pfam; PF00520; Ion_trans; 1.
CC Pfam; PF02214; K_tetra; 1.
CC PRINTS; PR00169; KCHANNEL.
CC PRINTS; PR01512; KV15CHANNEL.
CC PRINTS; PR01491; KVCHANNEL.
CC PRINTS; PR01496; SHAKERCHANNEL.
CC SMART; SM00225; BTB; 1.
KW Transport; Ion transport; Ionic channel; Voltage-gated channel;
KW Potassium channel; Potassium transport; Potassium; Transmembrane;
KW Glycoprotein; Phosphorylation; Multigene family.
FT TRANSMEM 240 258 SEGMENT S1.
FT TRANSMEM 314 334 SEGMENT S2.
FT TRANSMEM 345 366 SEGMENT S3.
FT TRANSMEM 386 407 SEGMENT S4.
FT TRANSMEM 422 443 SEGMENT S5.
FT TRANSMEM 483 504 SEGMENT S6.
FT DOMAIN 370 375 POLY-GLY.
FT CARBOHYD 114 114 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 179 179 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 288 288 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 601 AA; 65889 MW; 142DB9F8CBB43FE0 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 601;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 ANGSGP 18
Db 287 ANGSGP 292

RESULT 15
DD17 HUMAN STANDARD; PRT; 650 AA.
ID DD17 HUMAN STANDARD; PRT; 650 AA.
AC Q92841;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable RNA-dependent helicase p72 (DEAD-box protein p72) (DEAD-box
DE protein 17).
GN DDX17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
EX MEDLINE=97025363; PubMed=8871553;
RA Lamm G.M., Nicol S.M., Fuller-Pace F.V., Lamond A.I.;
RT "p72: a human nuclear DEAD box protein highly related to p68.";
RL Nucleic Acids Res. 24:3739-3747(1996).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruekewich R., Beare D.M.,
RA Clump M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,
RA Burrill W.D., Burton J., Carter C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Grahham D.W., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kerahaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leverisha M.A., Lloyd C., Lloyd D.M.,
RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spragon I., Stewart C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hua A., Kenton S., Lai H., Leo H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chiesoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA Scheet P., Walker C., Willems A., Wohlmann P., Pepin K., Nelson J.,
RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
RA McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
RA Kim U.J., Shizuya H., Simon M.I., Dumanaki J.P., Feyrard M., Kedra D.,
RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,
RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tilahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: RNA-DEPENDENT APYASE ACTIVITY.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DDX5/DDX17

```

```

CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U59321; AAC50787.1; -
CC EMBL; Z97056; CAB03792.1; -
CC EMBL; BC000595; AAH00595.1; -
CC PIR; S72367; S72367.
CC HSPP; Q58083; LHV8.
CC Genew; HGNC:2740; DDX17.
CC GO; GO:0005634; C:nucleus; TAS.
CC GO; GO:0008186; F:RNA dependent adenosinetriphosphatase activity; TAS.
CC GO; GO:0003724; F:RNA helicase activity; TAS.
CC GO; GO:0006396; P:RNA processing; TAS.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR000629; DEAD_box.
CC InterPro; IPR001650; Helicase_C.
CC Pfam; PF00270; DEAD; 1.
CC Pfam; PF00271; helicase_C; 1.
CC SMART; SM00487; DEXDC; 1.
CC SMART; SM00490; HELICG; 1.
CC PROSITE; PS00039; DEAD_ATP_HELICASE; 1.
CC ATP-binding; RNA-binding; Helicase; Nuclear protein.
CC DOMAIN 22 25 POLY-GLY.
CC NP BIND 136 143 ATP (POTENTIAL).
CC SITE 246 249 DEAD BOX.
CC FT SITE 477 484 POLY-GLY.
CC FT DOMAIN 539 647 POLY-PRO.
CC SQ SEQUENCE 650 AA; 72371 MW; E58AA249D23F66F3 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 650;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 YANGSG 17
Db 534 YANGSG 539

```

Search completed: November 21, 2003, 20:59:55
Job time : 5.85 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:45:57 ; Search time 24.4 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-77

Perfect score: 20

Sequence: 1 TDFDQGWGPISYANGSGPDQ 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertibrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	123	12	Q81457 hepatitis c
2	20	100.0	123	12	Q81449 hepatitis c
3	20	100.0	123	12	Q81444 hepatitis c
4	20	100.0	123	12	Q81442 hepatitis c
5	20	100.0	130	12	Q912L4 hepatitis c
6	20	100.0	130	12	Q912L6 hepatitis c
7	20	100.0	130	12	Q912L3 hepatitis c
8	20	100.0	130	12	Q912L5 hepatitis c
9	20	100.0	135	12	Q68196 hepatitis c
10	20	100.0	169	12	Q917R4 hepatitis c
11	20	100.0	2436	12	Q81756 hepatitis c
12	20	100.0	3011	12	Q91FE5 hepatitis c
13	19	95.0	135	12	Q68203 hepatitis c
14	19	95.0	137	12	Q9YK71 hepatitis c
15	19	95.0	137	12	Q9YIE2 hepatitis c
16	19	95.0	137	12	Q9YK72 hepatitis c

17	19	95.0	137	12	Q9YK68	Q9YK68 hepatitis c
18	19	95.0	137	12	Q9YK69	Q9YK69 hepatitis c
19	19	95.0	137	12	Q9YK70	Q9YK70 hepatitis c
20	19	95.0	139	12	Q91958	Q91958 hepatitis c
21	19	95.0	139	12	Q9YK74	Q9YK74 hepatitis c
22	19	95.0	139	12	Q9YK76	Q9YK76 hepatitis c
23	19	95.0	139	12	Q9YK73	Q9YK73 hepatitis c
24	19	95.0	139	12	Q91957	Q91957 hepatitis c
25	19	95.0	139	12	Q9YK77	Q9YK77 hepatitis c
26	19	95.0	139	12	Q9YJN1	Q9YJN1 hepatitis c
27	19	95.0	139	12	Q9YK75	Q9YK75 hepatitis c
28	19	95.0	326	12	Q9WG42	Q9WG42 hepatitis c
29	19	95.0	326	12	Q9WG40	Q9WG40 hepatitis c
30	19	95.0	326	12	Q9WG41	Q9WG41 hepatitis c
31	19	95.0	488	12	Q8B0R5	Q8B0R5 hepatitis c
32	19	95.0	488	12	Q8B0R4	Q8B0R4 hepatitis c
33	19	95.0	488	12	Q8B0R3	Q8B0R3 hepatitis c
34	19	95.0	778	12	Q04185	Q04185 hepatitis c
35	18	90.0	137	12	Q9YK67	Q9YK67 hepatitis c
36	18	90.0	137	12	Q91960	Q91960 hepatitis c
37	18	90.0	191	12	Q8JPB8	Q8JPB8 hepatitis c
38	17	85.0	123	12	Q81446	Q81446 hepatitis c
39	17	85.0	123	12	Q81456	Q81456 hepatitis c
40	17	85.0	151	12	Q9YK44	Q9YK44 hepatitis c
41	17	85.0	151	12	Q91981	Q91981 hepatitis c
42	17	85.0	151	12	Q9YK39	Q9YK39 hepatitis c
43	17	85.0	151	12	Q9YK64	Q9YK64 hepatitis c
44	17	85.0	151	12	Q9YK66	Q9YK66 hepatitis c
45	17	85.0	151	12	Q91976	Q91976 hepatitis c

ALIGNMENTS

RESULT 1

ID Q81457 o PRELIMINARY; PRT; 123 AA.
AC Q81457;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TremBLrel. 22, Last annotation update)
DE Unknown protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TWai-3;
RA Hall W.W., Suzuki H.;
RT "Genetic heterogeneity of Hepatitis C viruses in a population of HIV infected intravenous drug abusers (IVDAE).";
RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L19384; AAA45604.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 123
SQ SEQUENCE 123 AA; 13277 MW; 9C84646051F32118 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 123;

Best Local Similarity 100.0%; Pred. No. 9.7e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFDQGWGPISYANGSGPDQ 20

|||||

95 TDFDQGWGPISYANGSGPDQ 114

```

RESULT 2
Q81449 ID Q81449 PRELIMINARY; PRT; 123 AA.
AC Q81449;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE unknown protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCI-2;
RA Hall W.W., Suzuki H.;
RT "Genetic heterogeneity of Hepatitis C viruses in a population of HIV
RT infected intravenous drug abusers (IVDAs).";
RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L19378; AAA45599.1; -.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 123
SQ SEQUENCE 123 AA; 13305 MW; 7AA47DB834F1B419 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 123;
Best Local Similarity 100.0%; Pred. No. 9.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFDQGWGPISYANGSPDQ 20
DB 95 TDFDQGWGPISYANGSPDQ 114

RESULT 3
Q81444 ID Q81444 PRELIMINARY; PRT; 123 AA.
AC Q81444;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE unknown protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CR1-4;
RA Hall W.W., Suzuki H.;
RT "Genetic heterogeneity of Hepatitis C viruses in a population of HIV
RT infected intravenous drug abusers (IVDAs).";
RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L19373; AAA45594.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 123
SQ SEQUENCE 123 AA; 13231 MW; 47ABB2A628616ACD CRC64;

Query Match 100.0%; Score 20; DB 12; Length 123;

```

```

Best Local Similarity 100.0%; Pred. No. 9.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFDQGWGPISYANGSPDQ 20
DB 95 TDFDQGWGPISYANGSPDQ 114

RESULT 4
Q81442 ID Q81442 PRELIMINARY; PRT; 123 AA.
AC Q81442;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE unknown protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBI-5;
RA Hall W.W., Suzuki H.;
RT "Genetic heterogeneity of Hepatitis C viruses in a population of HIV
RT infected intravenous drug abusers (IVDAs).";
RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L19371; AAA45592.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 123
SQ SEQUENCE 123 AA; 13206 MW; 309F26F39ECFDBB2 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 123;
Best Local Similarity 100.0%; Pred. No. 9.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFDQGWGPISYANGSPDQ 20
DB 95 TDFDQGWGPISYANGSPDQ 114

RESULT 5
Q91ZL4 ID Q91ZL4 PRELIMINARY; PRT; 130 AA.
AC Q91ZL4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X304C-R;
RX MEDLINE=20173694; PubMed=10708420;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RA Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
RT replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL; AF230458; AAF68726.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.

```

KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 130 130
 SQ SEQUENCE 130 AA; 13855 MW; 0B693DB4E331F55D CRC64;
 Query Match 100.0%; Score 20; DB 12; Length 130;
 Best Local Similarity 100.0%; Pred. No. 1e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDFDQGWGPISYANGSGPDQ 20
 |||||
 DB 91 TDFDQGWGPISYANGSGPDQ 110
 RESULT 6
 Q91ZL6 PRELIMINARY; PRT; 130 AA.
 ID Q91ZL6;
 AC Q91ZL6;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=111103;
 RN [1]_TaxID=111103;
 RP SEQUENCE FROM N.A.
 RC STRAIN=X304C-I;
 RX MEDLINE=20173694; PubMed=10708420;
 RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
 Wang Y.M., Thomas D.L.;
 RT "Hypervariable region 1 sequence stability during hepatitis C virus
 replication in chimpanzees."
 RL J. Virol. 74:3058-3066(2000).
 DR EMBL; AF230456; AAF68724.1; -.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 130 130
 SQ SEQUENCE 130 AA; 13899 MW; 0B6A41E9E331F75F CRC64;
 Query Match 100.0%; Score 20; DB 12; Length 130;
 Best Local Similarity 100.0%; Pred. No. 1e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDFDQGWGPISYANGSGPDQ 20
 |||||
 DB 91 TDFDQGWGPISYANGSGPDQ 110
 RESULT 7
 Q91ZL3 PRELIMINARY; PRT; 130 AA.
 ID Q91ZL3;
 AC Q91ZL3;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=111103;
 RN [1]_TaxID=111103;
 RP SEQUENCE FROM N.A.
 RC STRAIN=X304C-U;
 RX MEDLINE=20173694; PubMed=10708420;
 RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
 Wang Y.M., Thomas D.L.;

RT "Hypervariable region 1 sequence stability during hepatitis C virus
 replication in chimpanzees."
 RL J. Virol. 74:3058-3066(2000).
 DR EMBL; AF230459; AAF68727.1; -.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 130 130
 SQ SEQUENCE 130 AA; 13827 MW; 0B692289E331F55D CRC64;
 Query Match 100.0%; Score 20; DB 12; Length 130;
 Best Local Similarity 100.0%; Pred. No. 1e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDFDQGWGPISYANGSGPDQ 20
 |||||
 DB 91 TDFDQGWGPISYANGSGPDQ 110
 RESULT 8
 Q91ZL5 PRELIMINARY; PRT; 130 AA.
 ID Q91ZL5;
 AC Q91ZL5;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=111103;
 RN [1]_TaxID=111103;
 RP SEQUENCE FROM N.A.
 RC STRAIN=X304C-K;
 RX MEDLINE=20173694; PubMed=10708420;
 RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
 Wang Y.M., Thomas D.L.;
 RT "Hypervariable region 1 sequence stability during hepatitis C virus
 replication in chimpanzees."
 RL J. Virol. 74:3058-3066(2000).
 DR EMBL; AF230457; AAF68725.1; -.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 130 130
 SQ SEQUENCE 130 AA; 13827 MW; 0B692289E331F55D CRC64;
 Query Match 100.0%; Score 20; DB 12; Length 130;
 Best Local Similarity 100.0%; Pred. No. 1e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDFDQGWGPISYANGSGPDQ 20
 |||||
 DB 91 TDFDQGWGPISYANGSGPDQ 110
 RESULT 9
 Q88196 PRELIMINARY; PRT; 135 AA.
 ID Q88196
 AC Q88196;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Envelope protein (Genome polyprotein) (Fragment).
 GN E2/NS1.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

```

OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1a;
RX MEDLINE=95146953; PubMed=7844535;
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT "Prevalence of hepatitis C virus sequence variants in South-East
RT Asia."; Virol. 76:211-215 (1995).
RL J. Gen. Virol. 76:211-215 (1995).
DR EMBL; U14232; AAC53921.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 135
SQ SEQUENCE 135 AA; 14408 MW; AD9A1939B546B8D1 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 135;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
Db 114 TDFDQGWGPISYANGSGPDQ 133

RESULT 10
Q917R4 PRELIMINARY; PRT; 169 AA.
AC Q917R4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HVR442-2;
RA Feucht H.;
RT "Long-term evolution of the hepatitis C virus hypervariable region.";
RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF416530; AAL10430.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 169
SQ SEQUENCE 169 AA; 18512 MW; C413AF2A2549A59B CRC64;

Query Match 100.0%; Score 20; DB 12; Length 169;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
Db 142 TDFDQGWGPISYANGSGPDQ 161

RESULT 11
Q91756 PRELIMINARY; PRT; 2436 AA.
ID Q91756

```

```

AC Q81756;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC Choo Q.-L., Richman K., Han J.;
RT "The nucleotide sequence of the Hepatitis C viral genome.";
RT Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.
DR EMBL; M32084; AAA45677.1; -.
DR HSP; P27958; IAIV.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDc; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05052; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydroxylase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
FT NON_TER 1
FT NON_TER 2436
SQ SEQUENCE 2436 AA; 264734 MW; D7B9872900BE3125 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 2436;
Best Local Similarity 100.0%; Pred. No. 1.3e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDFDQGWGPISYANGSGPDQ 20
Db 13 TDFDQGWGPISYANGSGPDQ 32

RESULT 12
Q91FES PRELIMINARY; PRT; 3011 AA.
ID Q91FES
AC Q91FES;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
RT sequence.";

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:22 ; Search time 32.1 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-78

Perfect score: 20

Sequence: 1 ANGSGPDRPYCWHYPPKPC 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A Geneseq 19Jun03.*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	16 AAR84507	Hepatitis C virus
2	20	100.0	20	17 AAR91001	HCV E2 peptide E2-
3	20	100.0	20	23 AAO18707	Hepatitis C virus
4	20	100.0	42	13 AAR20751	Peptide 1 based on
5	20	100.0	76	21 AAB18528	Protein encoded by
6	20	100.0	144	14 AAR33998	HC-J1 E2/NS1 prote
7	20	100.0	174	14 AAR34441	Sequence of glycop
8	20	100.0	192	19 AAW67009	HCV nucleocapsid c
9	20	100.0	254	22 AAB68043	Amino acid sequenc

10	20	100.0	278	14 AAR33997	Th E2/NS1 protein.
11	20	100.0	319	15 AAR45330	Anti-HCV antibody
12	20	100.0	363	22 AAB68042	Amino acid sequenc
13	20	100.0	402	14 AAR34438	Sequence of glycop
14	20	100.0	454	10 AAP90183	Sequence of hepati
15	20	100.0	454	10 AAP92049	Sequence encoded b
16	20	100.0	454	21 AAB18528	Protein encoded by
17	20	100.0	463	14 AAR33588	HCV CKS-NS1 fusi
18	20	100.0	463	14 AAR33187	Sequence of subfra
19	20	100.0	463	22 AAB69007	HCV recombinant an
20	20	100.0	480	14 AAR33992	HCV-1 E2/NS1 prote
21	20	100.0	513	13 AAR24086	NANB hepatitis vir
22	20	100.0	531	22 AAE02622	Chimeric HCV E2661
23	20	100.0	621	14 AAR33185	Sequence of subfra
24	20	100.0	622	14 AAR33591	HCV CKS-NS1-NS1S
25	20	100.0	622	22 AAB69010	HCV recombinant an
26	20	100.0	637	24 AAB57410	Hepatitis C virus
27	20	100.0	663	17 AAR29355	HCV1 E2 + NS2 poly
28	20	100.0	663	20 AAW67615	Hepatitis C virus
29	20	100.0	733	14 AAR38278	NANB hepatitis vir
30	20	100.0	738	14 AAR33592	HCV CKS-full lengt
31	20	100.0	738	22 AAB69011	HCV recombinant an
32	20	100.0	2435	13 AAR25135	HCV polypeptide 1.
33	20	100.0	2436	10 AAP92050	Sequence encoded i
34	20	100.0	2436	10 AAP90288	Peptide encoded by
35	20	100.0	2436	13 AAR28582	HCV amino acid seq
36	20	100.0	2772	11 AAR08123	Hepatitis C virus
37	20	100.0	2772	21 AAB18540	Protein encoded by
38	20	100.0	2816	14 AAR34009	HCV-1 polyprotein.
39	20	100.0	2894	13 AAR24440	Composite HCV HC-J
40	20	100.0	2894	16 AAR70230	Composite hepatitis
41	20	100.0	2955	11 AAR08124	Hepatitis C virus
42	20	100.0	2955	20 AAY14975	Amino acid sequenc
43	20	100.0	2955	21 AAB18541	Polyprotein encode
44	20	100.0	3011	13 AAR21519	Compiled HCV sequ
45	20	100.0	3011	14 AAR31621	Hepatitis C virus

ALIGNMENTS

RESULT 1 0
AAR84507
ID AAR84507 standard; peptide; 20 AA.
AC AAR84507;
XX
XX
DT 06-JAN-1997 (first entry)
XX
DE Hepatitis C virus peptide NS1-5 (residues 475-494).
XX
DE Hepatitis C virus; HCV; immunogen; non-structural region; NS1;
KW immunodominant; T cell epitope; vaccine.
XX
OS Hepatitis C virus.
XX
PN WO9512677-A2.
XX
PD 11-MAY-1995.
XX
PF 28-OCT-1994; 94WO-EP03555.
XX
PR 04-NOV-1993; 93EP-0402718.
XX
PA (INNO-) INNOGENETICS NV.
XX
PI Deleys R, Leroux-Roels G, Maertens G;
XX
DR WPI, 1995-193822/25.
XX
PT Hepatitis C virus immunogenic polypeptide contg. a T-cell
PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in
PT production of vaccines, therapeutic agents, etc.

```

XX Example 4; Page 51; 105pp; English.
PS A series of overlapping peptides (including the present sequence) was
CC synthesised based on sequences in the core, E1 and E2/NS1 regions of
CC hepatitis C virus. The peptides were used as antigens in lympho-
CC proliferative assays to identify the main T-cell epitopes.
XX Sequence 20 AA;
SQ Query Match 100.0%; Score 20; DB 16; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ANGSGPDQRPYCWHPKPC 20
Db 1 ANGSGPDQRPYCWHPKPC 20
RESULT 2
AAR91001
ID AAR91001 standard; peptide; 20 AA.
AC AAR91001;
XX 25-SEP-1996 (first entry)
DT HCV E2 peptide E2-5B for competition studies.
DE HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;
KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.
XX Synthetic.
OS WO9604385-A2.
XX 15-FEB-1996.
PD 31-JUL-1995; 95WO-EP03031.
PF 29-JUL-1994; 94EP-0870132.
PR (INNO-) INNOGENETICS NV.
PA Bosman F, Buyse M, De Martynoff G, Maertens G;
XX WPI; 1996-129401/13.
DR Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope
XX proteins - in presence of disulphide bond cleavage agent, to
PT produce proteins suitable for direct use in vaccines or diagnostic
PT assays of HCV
XX Example 7; Page 67; 146pp; English.
PS AAR90965-R30980 and AAR90995-R31015 represent synthetic hepatitis C
XX virus (HCV) E1 and E2 peptides used in competition studies. This
CC sequence represents a synthetic E2 peptide, and corresponds to residues
CC 475-494 of the E2 protein sequence. These sequences are useful for in
CC vitro monitoring of HCV disease, or prognosis of the response to
CC interferon treatment of patients suffering from HCV infection. These
CC sequences compete with the proteins produced by AAT12704-T12709 and
CC AAT12961-T12974, which are included in vectors for the production of
CC recombinant E1, E2, and E1/E2. The recombinant proteins can then be
CC isolated and purified by carrying out a disulphide bond cleavage, or a
CC reduction step with a disulphide bond cleavage agent, after lysis of
CC recombinant host cells. The constructs containing the purified HCV
CC envelope proteins can be used for vaccinating humans against HCV, for in
CC vitro detection of HCV antibodies in a sample, and in a serotyping assay
CC for detecting one or more serological types of HCV present in a
CC biological sample. The constructs can also be immobilised on a solid
CC substrate and incorporated into a reversed phase hybridisation assay for
CC determining the presence or the genotype of HCV. The new purification

```

```

CC method preserves the conformation of the recombinantly expressed E1, E2
CC and E1/E2, and eliminates contaminating proteins. Antigens isolated
CC using this method are more reactive with human sera than those isolated
CC by known techniques.
XX Sequence 20 AA;
SQ Query Match 100.0%; Score 20; DB 17; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ANGSGPDQRPYCWHPKPC 20
Db 1 ANGSGPDQRPYCWHPKPC 20
RESULT 3
AAR18707
ID AAR18707 standard; Peptide; 20 AA.
XX AAR18707;
AC AAR18707;
XX 24-OCT-2002 (first entry)
DT Hepatitis C virus E2 protein derived peptide E2-5B.
DE Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide;
KW immunostimulant; vaccine.
XX Hepatitis C virus.
OS WO200255548-A2.
XX 18-JUL-2002.
PD 11-JAN-2002; 2002WO-EP00219.
PF 11-JAN-2001; 2001US-260699P.
PR 30-AUG-2001; 2001US-315768P.
XX (INNO-) INNOGENETICS NV.
PA Maertens G, Bosman F, Buyse M;
XX WPI; 2002-599657/64.
DR New therapeutic vaccine compositions comprising at least one purified
XX recombinant hepatitis C virus (HCV) single or specific oligomeric
PT recombinant envelope protein E1 or E2, useful for immunizing humans
PT from HCV infection
XX Example 7; Page 227; 243pp; English.
PS The present invention relates to new therapeutic vaccine compositions for
XX inducing hepatitis C virus (HCV)-specific antibodies, comprising a
CC composition containing at least one purified recombinant HCV single or
CC specific oligomeric recombinant envelope proteins selected from an E1 and
CC an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are
CC useful for inducing HCV-specific antibodies or for immunising humans
CC against HCV. The recombinant HCV E1 and/or E2 proteins are useful as
CC vaccine or therapeutics, in HCV screening and confirmatory antibody
CC tests, for raising antibodies, in the preparation of medicament, and for
CC in vitro monitoring of HCV disease or prognosing the response to
CC treatment of patients suffering from HCV infection. The present sequence
CC is a peptide derived from the proteins of the invention.
XX Sequence 20 AA;
SQ Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ANGSGPDQRPYCWHPKPC 20

```

Db |||||||||||||||||||
1 ANGSGPDQRPYCWHPKPC 20

RESULT 4

AAR20751
ID AAR20751 standard; peptide; 42 AA.

XX AAR20751;
XX AC
XX 25-MAR-2003 (updated)
DT 05-MAY-1992 (first entry)
XX
XX Peptide 1 based on immunoreactive region of Hepatitis C virus.
DE
XX
XX Non-A, non-B hepatitis virus; non-structural protein; vaccine.
KW
XX
XX Synthetic.
OS
XX
XX EP468527-A.
FN
XX
XX 29-JAN-1992.
PD
XX
XX 26-JUL-1991; 91EP-0112620.
PF
XX
XX 24-JUN-1991; 91US-0719819.
PR
XX 26-JUL-1990; 90US-0558799.
PR
XX 07-FEB-1991; 91US-0651735.
PR
XX 11-MAR-1991; 91US-0667275.
XX
XX (UNBI-) UNITED BIOMEDICAL INC.
PA
XX
XX Chang YW, Hosein B;
PI
XX
XX WPI; 1992-034279/05.
DR
XX
XX New synthetic peptide specific for HCV antibodies - for detection
PT of HCV or NANBHV e.g. by enzyme-linked immunosorbent assay and is
PT immunogen for preparation of vaccines
XX
XX Claim 1; Page 88; 98pp; English.
PS
XX

This peptide is one of 19 specifically claimed antigens based on the immunoreactive regions of the envelope protein and non-structural proteins NS-1, NS-2, NS-3 and NS-5 for the Hepatitis C virus. The invention also concerns analogues, segments, mixtures, conjugates and polymers of these peptides. The C-terminal amino acid may be amidated. See also AAR20752-R20782.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
XX Sequence 42 AA;

Query Match 100.0%; Score 20; DB 13; Length 42;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20

Db 9 ANGSGPDQRPYCWHPKPC 28

RESULT 5

AAB18528
ID AAB18528 standard; Protein; 76 AA.

XX AAB18528;
AC
XX 15-JAN-2001 (first entry)
DT
XX
XX Protein encoded by a novel hepatitis C virus cDNA clone 26j.
DE
XX
XX Hepatitis C virus; HCV; antisense polynucleotide; polypeptide; viral infectivity; viral replication.
KW

XX Hepatitis C virus.
OS
XX
XX EP1034785-A2.
PN
XX
XX 13-SEP-2000.
PD
XX
XX 16-MAR-1990; 2000EP-0109602.
PF
XX
XX 17-MAR-1989; 89US-0325338.
PR
XX 20-APR-1989; 89US-0341334.
PR
XX 18-MAY-1989; 89US-0355002.
PR
XX 16-MAR-1990; 90EP-0302866.
PR
XX
XX (CHIR) CHIRON CORP.
PA
XX
XX Houghton M, Choo Q, Kuo G;
FI
XX
XX WPI; 2000-566891/53.
FN
XX
XX N-PSDB; AAA75284.
DR
XX
XX Novel composition comprising a hepatitis C virus antisense
PT polynucleotide which is complementary to or corresponds to a sense
PT strand of the virus genome, and selectively hybridises to it -
XX
XX Example; Fig 4; 75pp; English.

The specification describes a pharmaceutical composition which comprises a hepatitis C virus (HCV) antisense polynucleotide. The HCV is characterized by a positive stranded RNA genome which has 40% homology at the polypeptide level to a HCV polypeptide. The antisense polynucleotide binds to cellular polynucleotides which enhance and/or are required for viral infectivity, replicative ability or chronicity. The antisense polynucleotides may also be designed to bind with high specificity, to be of increased stability, to be stable and to have low toxicity. The composition also comprises an agent which causes viral RNA to be inactive. The composition is used for preventing HCV replication in a system. The present sequence is encoded by a novel HCV cDNA sequence, which is used in the course of the invention.

XX Sequence 76 AA;

Query Match 100.0%; Score 20; DB 21; Length 76;
Best Local Similarity 100.0%; Pred. No. 2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20

Db 35 ANGSGPDQRPYCWHPKPC 54

RESULT 6
AAR33998
ID AAR33998 standard; Protein; 144 AA.

XX AAR33998;
AC

XX 25-MAR-2003 (updated)
DT 26-JUL-1993 (first entry)
XX
XX
XX HC-J1 E2/NS1 protein.

XX Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
KW domain; immunological; cross-reactive; envelope protein; vaccine;
KW gp53 (BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.

XX Synthetic.

XX WO9306126-A1.

XX 01-APR-1993.
PD

```

XX 11-SEP-1992; 92WO-US07683.
XX
XX
PR 13-SEP-1991; 91US-0759575.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Houghton M, Weiner AJ;
XX
XX WPI; 1993-117468/14.
XX
XX Immuno-reactive hepatitis C virus polypeptide compsns. - contg.
PT at least 2 sequences from the first variable domain of distinct
PT HCV isolates
XX
XX Disclosure; Fig 3; 106pp; English.
XX
XX The sequences given in AAR33992-002 represent a portion of the E2/NS1
CC protein encoded by group I and group II HCV isolates, from amino acid
CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
CC 30 amino acids which shows large variation between nearly all isolates.
CC This is an important immunoreactive domain. This putative envelope
CC glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera
CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
CC flaviviruses, both of which confer protective immunity in hosts
CC vaccinated with these polypeptides. It has been discovered that a
CC number of important HCV epitopes vary among viral isolates and that
CC these epitopes can be mapped to specific domains. This meant that
CC immunologically cross-reactive polypeptides which focus on variable
CC rather than constant domains can be produced. See also AAR39134-48
CC and AAR33982-91.
XX (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 144 AA;
SQ
Query Match 100.0%; Score 20; DB 14; Length 144;
Best Local Similarity 100.0%; Pred. No. 3.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ANGSGPDQPCYCHWYPPKPC 20
DB 106 ANGSGPDQPCYCHWYPPKPC 125
RESULT 7
AAR34441
ID AAR34441 standard; Protein; 174 AA.
XX
XX AAR34441;
XX
XX 25-MAR-2003 (updated)
DT 09-AUG-1993 (first entry)
XX
XX Sequence of glycoprotein E2/NS1 in clone J1 (JM).
XX Hepatitis C virus; envelope protein; glycoprotein; E2/NS1;
XX diagnostic reagent.
XX Hepatitis C virus.
XX EP537626-A1.
XX
XX 21-APR-1993.
PD
XX 08-OCT-1992; 92EP-0117191.
PF
XX 08-OCT-1991; 91JP-0260824.
PR
XX (NAHE-) NAT INST OF HEALTH.
XX
XX Harada S, Honda Y, Miyamura T, Saito I;
XX

```

```

DR WPI; 1993-127516/16.
XX N-PSDB; AAR40333.
XX
XX Diagnostic reagent for hepatitis C virus - comprises second
PT envelope protein or first non-structural protein encoded by HCV
PT gene and has sugar chain
XX
XX Claim 2; Pages 40-41; 58pp; English.
XX
XX Glycoprotein E2/NS1 is derived from the second envelope protein or
CC first non-structural protein encoded by the genome of HCV. The
CC nucleic acid is extracted from the serum of the patient of hepatitis
CC C. The serum is pref. mixed with transfer RNA (tRNA) as a carrier
CC of virus RNA. As a technique of cloning cDNA from the nucleic acid,
CC it is preferred to use polymerase chain reaction method. In the
CC reaction, any commercially available random primers or synthesized
CC DNA having a base sequence similar to that of primer AS1 may be used
CC as a primer. Representative examples of sense primers include S1.
XX (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 174 AA;
SQ
Query Match 100.0%; Score 20; DB 14; Length 174;
Best Local Similarity 100.0%; Pred. No. 3.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ANGSGPDQPCYCHWYPPKPC 20
DB 136 ANGSGPDQPCYCHWYPPKPC 155
RESULT 8
AAR67009
ID AAR67009 standard; protein; 192 AA.
XX
XX AAR67009;
XX
XX 02-MAR-1999 (first entry)
DT
XX HCV nucleocapsid core protein.
DE
XX Hepatitis C virus; HCV; nucleocapsid; core protein; analogue; antibody;
XX non-structural protein; thioamide bond; peptide bond.
KW
XX Hepatitis C virus.
OS
XX Key Location/Qualifiers
FH Misc-difference 79
FT /note= "given in specification as Try"
FT
XX
XX JPI0226698-A.
PN
XX 25-AUG-1998.
PD
XX 19-FEB-1997; 97JP-0034702.
PF
XX 19-FEB-1997; 97JP-0034702.
PR
XX (KYOW ) KYOWA MEDEX KK.
PA
XX WPI; 1998-515103/44.
XX
XX Determination of antibody in sample - uses peptide analog absorbed
PT or chemically bound on carrier as antigen
PT
XX Disclosure; Page 4; 13pp; Japanese.
XX
XX This sequence represents the Hepatitis C virus (HCV) nucleocapsid core
CC protein. The invention relates to peptide analogues derived from HCV
CC proteins, e.g. AAR67417-W67426, which can be used for the determination
CC of anti-HCV antibodies in a sample. Preferably the peptide analogues
CC contain one or more thioamide peptide bonds where at least one oxygen
CC atom of the peptide bond is replaced by sulphur atom. The peptide

```

CC analogues can be adsorbed or chemically bound to a carrier.

SQ Sequence 192 AA;

Query Match 100.0%; Score 20; DB 19; Length 192;
Best Local Similarity 100.0%; Pred. No. 4.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANGSGPDQRPYCWHPKPC 20

|||||
25 ANGSGPDQRPYCWHPKPC 44

RESULT 9

AAB68043

ID AAB68043 standard; protein; 254 AA.

XX AC AAB68043;

XX DT 29-JUN-2001 (first entry)

XX DE Amino acid sequence of water soluble variant of envelope E2 protein.

XX KW Eo protein; HCV; envelope protein; E2 protein; HCV infection;

XX KW HCV attachment.

XX OS Synthetic.

XX OS Hepatitis C virus.

XX PN WO200122984-A1.

XX PD 05-APR-2001.

XX PF 26-SEP-2000; 2000WO-US26395.

XX PR 29-SEP-1999; 99US-0407430.

XX PA (UYCO) UNIV COLUMBIA NEW YORK.

XX PI Worman HJ, Mamiya N;

XX DR WPI; 2001-273486/28.

XX PT Treating or preventing hepatitis C virus infection in a subject,
XX PT involves administering hepatitis C virus envelope protein E2 binding
XX PT agents -

XX PS Claim 5; Fig 8; 46pp; English.

XX CC The present sequence represents a water soluble variant of a Hepatitis C
XX CC virus (HCV) envelope E2 protein. Agents that bind to the HCV E2 protein
XX CC (such as the human Eo protein), and so inhibit the attachment of HCV onto
XX CC cells (especially liver cells), are used to treat HCV infections in
XX CC mammals, in particular humans. The specification also describes a method
XX CC for identifying a compound which can be used for treating or preventing
XX CC HCV in a subject and which can inhibit the attachment of HCV onto cells
XX CC by inhibiting the binding of HCV envelope E2 protein to a cellular
XX CC protein associated with HCV attachment and entry into cells. The method
XX CC comprises incubating the compound, HCV envelope E2 protein or its variant
XX CC and a cellular protein capable of specifically binding to the HCV E2
XX CC protein under suitable reaction conditions; determining the interactions
XX CC between HCV envelope E2 protein and cellular protein in the presence and
XX CC absence of the compound; and comparing the interaction to identify a
XX CC compound which can inhibit the attachment of HCV onto cells.

SQ Sequence 254 AA;

Query Match 100.0%; Score 20; DB 22; Length 254;
Best Local Similarity 100.0%; Pred. No. 5.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANGSGPDQRPYCWHPKPC 20

|||||

Db 70 ANGSGPDQRPYCWHPKPC 89

RESULT 10

AAR33997

ID AAR33997 standard; Protein; 278 AA.

XX AC AAR33997;

XX DT 25-MAR-2003 (updated)

XX DT 26-JUL-1993 (first entry)

XX DE Th E2/NS1 protein.

XX KW Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
XX KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
XX KW domain; immunological; cross-reactive; envelope protein; vaccine;
XX KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.

XX OS Synthetic.

XX PN WO9306126-A1.

XX PD 01-APR-1993.

XX PF 11-SEP-1992; '92WO-US07683.

XX PR 13-SEP-1991; 91US-0759575.

XX PA (CHIR) CHIRON CORP.

XX PI Houghton M, Weiner AJ;

XX DR WPI; 1993-117468/14.

XX PT Immuno-reactive hepatitis C virus polypeptide compens. - contg.
XX PT at least 2 sequences from the first variable domain of distinct
XX PT HCV isolates

XX PS Disclosure; Fig 3; 106pp; English.

XX CC The sequences given in AAR33992-002 represent a portion of the E2/NS1
XX CC protein encoded by group I and group II HCV isolates, from amino acid
XX CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
XX CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
XX CC 30 amino acids which shows large variation between nearly all isolates.
XX CC This is an important immunoreactive domain. This putative envelope
XX CC glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera
XX CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
XX CC flaviviruses, both of which confer protective immunity in hosts
XX CC vaccinated with these polypeptides. It has been discovered that a
XX CC number of important HCV epitopes vary among viral isolates and that
XX CC these epitopes can be mapped to specific domains. This meant that
XX CC immunologically cross-reactive polypeptides which focus on variable
XX CC rather than constant domains can be produced. See also AAQ93134-48
XX CC and AAR33982-91.
XX CC (Updated on 25-MAR-2003 to correct PN field.)

SQ Sequence 278 AA;

Query Match 100.0%; Score 20; DB 14; Length 278;
Best Local Similarity 100.0%; Pred. No. 5.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANGSGPDQRPYCWHPKPC 20

|||||

Db 106 ANGSGPDQRPYCWHPKPC 125

RESULT 11

AAR45330

ID AAR45330 standard; protein; 319 AA.

XX

AC AAR45330;
 XX
 DT 28-JUN-1994 (first entry)
 XX
 DE Anti-HCV antibody reactive protein #1.
 XX
 KW Hepatitis C virus; HCV; envelope; region; anti-HCV; antibody; vaccine.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Peptide 39...69
 FT /label= EPI
 FT /note= "Claim 1"
 XX
 FN JP05310786-A.
 XX
 PD 22-NOV-1993.
 XX
 PF 30-APR-1992; 92JP-0111846.
 XX
 PR 30-APR-1992; 92JP-0111846.
 XX
 PA (OLYU) OLYMPUS OPTICAL CO LTD.
 XX
 DR WPI; 1994-002315/01.
 XX
 PT Peptide of HCV envelope region - having high reactivity against
 PT anti-hepatitis C virus antibody and having at least 6 amino acids
 XX
 PS Disclosure; Page 6-7; 11pp; Japanese.
 XX
 CC The sequences given in AAR45330-31 represent fragments derived from
 CC the hepatitis C virus (HCV) envelope region that are highly reactive
 CC against anti-HCV antibodies. These protein fragments may be used in
 CC vaccines against HCV. They have a high reactivity against anti-HCV
 CC antibodies regardless of the original HCV. They can be used to
 CC detect HCV infection but not the onset of infection.
 XX
 SQ Sequence 319 AA;
 Query Match 100.0%; Score 20; DB 15; Length 319;
 Best Local Similarity 100.0%; Pred. No. 6.1e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ANGSGPDQRPYCWHPKPC 20
 Db 284 ANGSGPDQRPYCWHPKPC 303
 RESULT 12
 AAR68042
 ID AAB68042 standard; protein; 363 AA.
 XX
 AC AAB68042;
 XX
 DT 29-JUN-2001 (first entry)
 XX
 DE Amino acid sequence of a Hepatitis C virus envelope E2 protein.
 XX
 KW Eo protein; HCV; envelope protein; E2 protein; HCV infection;
 KW HCV attachment.
 XX
 OS Hepatitis C virus.
 XX
 FN WO200122984-A1.
 XX
 PD 05-APR-2001.
 XX
 PF 26-SEP-2000; 2000WO-US26395.
 XX
 PR 29-SEP-1999; 99US-0407430.
 XX

PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 PI Worman HJ, Mamiya N;
 XX
 DR WPI; 2001-273486/28.
 XX
 FT Treating or preventing hepatitis C virus infection in a subject,
 FT involves administering hepatitis C virus envelope protein E2 binding
 FT agents -
 XX
 PS Claim 3; Fig 7; 46pp; English.
 XX
 CC The present sequence represents a Hepatitis C virus (HCV) envelope
 CC E2 protein. Agents that bind to the HCV E2 protein (such as the human
 CC Eo protein), and so inhibit the attachment of HCV onto cells
 CC (especially liver cells), are used to treat HCV infections in mammals,
 CC in particular humans. The specification also describes a method for
 CC identifying a compound which can be used for treating or preventing
 CC HCV in a subject and which can inhibit the attachment of HCV onto cells
 CC by inhibiting the binding of HCV envelope E2 protein to a cellular
 CC protein associated with HCV attachment and entry into cells. The method
 CC comprises incubating the compound, HCV envelope E2 protein or its variant
 CC and a cellular protein capable of specifically binding to the HCV E2
 CC protein under suitable reaction conditions; determining the interactions
 CC between HCV envelope E2 protein and cellular protein in the presence and
 CC absence of the compound; and comparing the interaction to identify a
 CC compound which can inhibit the attachment of HCV onto cells.
 XX
 SQ Sequence 363 AA;
 Query Match 100.0%; Score 20; DB 22; Length 363;
 Best Local Similarity 100.0%; Pred. No. 6.7e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ANGSGPDQRPYCWHPKPC 20
 Db 92 ANGSGPDQRPYCWHPKPC 111
 RESULT 13
 AAR34438
 ID AAR34438 standard; Protein; 402 AA.
 XX
 AC AAR34438;
 XX
 DT 25-MAR-2003 (updated)
 DT 09-AUG-1993 (first entry)
 XX
 DE Sequence of glycoprotein E2/NS1 in clone HCV1.
 XX
 KW Hepatitis C virus; envelope protein; glycoprotein; E2/NS1;
 KW diagnostic reagent.
 XX
 OS Hepatitis C virus.
 XX
 PN EP537626-A1.
 XX
 PD 21-APR-1993.
 XX
 PF 08-OCT-1992; 92EP-0117191.
 XX
 PR 08-OCT-1991; 91JP-0260824.
 XX
 PA (NAHE-) NAT INST OF HEALTH.
 XX
 PI Harada S, Honda Y, Miyamura T, Saito I;
 XX
 DR WPI; 1993-127516/16.
 DR N-PSDB; AAQ40330.
 XX
 XX Diagnostic reagent for hepatitis C virus - comprises second
 PT envelope protein or first non-structural protein encoded by HCV
 PT gene and has sugar chain

XX Claim 2; Pages 30-32; 58pp; English.
 XX Glycoprotein E2/NS1 is derived from the second envelope protein or
 CC first non-structural protein encoded by the genome of HCV. The
 CC nucleic acid is extracted from the serum of the patient of hepatitis
 CC C. The serum is pref. mixed with transfer RNA (tRNA) as a carrier
 CC of virus RNA. As a technique of cloning cDNA from the nucleic acid,
 CC it is preferred to use polymerase chain reaction method. In the
 CC reaction, any commercially available random primers or synthesized
 CC DNA having a base sequence similar to that of primer AS1 may be used
 CC as a primer. Representative examples of sense primers include S1.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX SQ Sequence 402 AA;
 XX Query Match 100.0%; Score 20; DB 14; Length 402;
 XX Best Local Similarity 100.0%; Pred. No. 7.2e-13;
 XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ANGSGPDQRPYCWHPKPC 20
 Db 136 ANGSGPDQRPYCWHPKPC 155
 ||||||||||||||||
 RESULT 14
 AAP90183
 ID AAP90183 standard; protein; 454 AA.
 XX AC AAP90183;
 XX DT 25-MAR-2003 (updated)
 XX DT 01-NOV-1989 (first entry)
 XX DE Sequence of hepatitis C virus cDNA insert in clone k9-1.
 XX KW Hepatitis C virus; clone k9-1; probe; vaccine.
 XX OS Pan troglodytes.
 XX FH Key Location/Qualifiers
 XX FT Region 97..454
 XX GB2212511-A.
 XX PD 26-JUL-1989.
 XX PF 18-NOV-1988; 88GB-0027024.
 XX PR 18-NOV-1987; 87US-0122714.
 XX PR 30-DEC-1987; 87US-0139886.
 XX PR 26-FEB-1988; 88US-0161072.
 XX PR 26-OCT-1988; 88US-0263584.
 XX PA (CHIR) CHIRON CORPORATION.
 XX PI Houghton M, Choo QL, Kuo G;
 XX WPI; 1989-215054/30.
 XX N-PSDB; AAN90335.
 XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
 PT polypeptide(s) and antibodies for diagnosis, prevention and
 PT treatment of infection.
 XX Disclosure; fig 46; 235pp; English.
 XX The sequence is the peptide encoded by the hepatitis C virus
 CC (HCV) cDNA insert in clone k9-1 (see AAN90335). The polypeptides
 CC are used to diagnose HCV-induced NANBH, to raise antibodies for
 CC immunoassay or treatment, or to produce vaccines.
 CC The region shown overlaps the cDNA of AAN90327.
 CC (Updated on 25-MAR-2003 to correct PR field.)

XX SQ Sequence 454 AA;
 XX Query Match 100.0%; Score 20; DB 10; Length 454;
 XX Best Local Similarity 100.0%; Pred. No. 8e-13;
 XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ANGSGPDQRPYCWHPKPC 20
 Db 25 ANGSGPDQRPYCWHPKPC 44
 ||||||||||||||||
 RESULT 15
 AAP92049
 ID AAP92049 standard; protein; 454 AA.
 XX AC AAP92049;
 XX DT 25-MAR-2003 (updated)
 XX DT 21-JAN-1991 (first entry)
 XX DE Sequence encoded by segment of the hepatitis C virus (HCV) cDNA sequence
 XX DE in clone K9-1.
 XX KW Non-a non-B hepatitis; probe; vaccine; diagnosis;
 XX KW passive immunotherapy; antigen.
 XX OS Hepatitis C virus.
 XX PN EP318216-A.
 XX PD 31-MAY-1989.
 XX PF 18-NOV-1988; 88EP-0310922.
 XX PR 18-NOV-1987; 87US-0122714.
 XX PR 30-DEC-1987; 87US-0139886.
 XX PR 26-FEB-1988; 88US-0161072.
 XX PR 06-MAY-1988; 88US-0191263.
 XX PR 26-OCT-1988; 88US-0263584.
 XX PR 14-NOV-1988; 88US-0271450.
 XX PA (CHIR) CHIRON CORP.
 XX PI Houghton M, Choo QL, Kuo G;
 XX WPI; 1989-159274/22.
 XX N-PSDB; AAN92105.
 XX Purified hepatitis C virus -
 PT and assorted nucleic acids and polypeptide(s)
 XX Example; Fig 46-1 - 46-2; 139pp; English.
 XX Purified hepatitis C virus (HCV) and purified or recombinant HCV nucleic
 CC acids (NAs), encoding HCV polynucleotides or epitopes, and polypeptides
 CC are claimed. HCV is a causative agent of non-A, non-B hepatitis (NANBH).
 CC The NAs may be used to design probes for detn. of HCV NAs in samples.
 CC The polypeptides may be used as immunoassay reagents and vaccines, and
 CC to produce antibodies useful for diagnosis and passive immunotherapy.
 CC The purified virus may also be used in vaccines.
 CC (Updated on 25-MAR-2003 to correct PR field.)
 CC (Updated on 25-MAR-2003 to correct PA field.)
 XX SQ Sequence 454 AA;
 XX Query Match 100.0%; Score 20; DB 10; Length 454;
 XX Best Local Similarity 100.0%; Pred. No. 8e-13;
 XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ANGSGPDQRPYCWHPKPC 20
 Db 25 ANGSGPDQRPYCWHPKPC 44
 ||||||||||||||||

Search completed: November 21, 2003, 20:58:04
Job time : 33.15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.6 Seconds
(without alignments)
79.832 Million cell updates/sec

Title: US-09-973-025-78

Perfect score: 20

Sequence: 1 ANGSGPDQRYCWHYPPKPC 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-08-612-973-78
2	20	100.0	20	3	US-08-927-597-78
3	20	100.0	20	4	US-08-635-886C-34
4	20	100.0	34	1	US-08-262-037-38
5	20	100.0	37	1	US-08-262-037-34
6	20	100.0	42	1	US-08-262-037-1
7	20	100.0	45	1	US-08-262-037-35
8	20	100.0	50	1	US-08-262-037-39
9	20	100.0	52	1	US-08-262-037-36
10	20	100.0	57	1	US-08-262-037-40
11	20	100.0	76	3	US-08-444-818-79
12	20	100.0	144	1	US-08-440-103-20
13	20	100.0	144	1	US-08-440-542-20
14	20	100.0	144	1	US-08-231-368-20
15	20	100.0	144	1	US-08-440-210-20
16	20	100.0	144	4	US-09-046-604-20
17	20	100.0	174	1	US-08-460-806-19
18	20	100.0	174	1	US-08-325-630-19
19	20	100.0	278	1	US-08-440-103-15
20	20	100.0	278	1	US-08-440-542-15
21	20	100.0	278	1	US-08-231-368-15
22	20	100.0	278	1	US-08-440-210-15
23	20	100.0	278	4	US-09-046-604-15
24	20	100.0	402	1	US-08-460-806-13
25	20	100.0	402	1	US-08-325-630-13
26	20	100.0	403	2	US-08-483-695-39
27	20	100.0	403	2	US-07-965-285-39

28	20	100.0	403	2	US-08-487-231-39	Sequence 39, Appl
29	20	100.0	403	3	US-09-201-912-39	Sequence 39, Appl
30	20	100.0	454	3	US-08-444-818-73	Sequence 73, Appl
31	20	100.0	463	1	US-07-748-292-9	Sequence 9, Appl
32	20	100.0	463	3	US-08-867-611-31	Sequence 31, Appl
33	20	100.0	463	5	PCT-US92-06965A-1	Sequence 1, Appl
34	20	100.0	480	1	US-08-440-103-14	Sequence 14, Appl
35	20	100.0	480	1	US-08-440-542-14	Sequence 14, Appl
36	20	100.0	480	1	US-08-231-368-14	Sequence 14, Appl
37	20	100.0	480	1	US-08-440-210-14	Sequence 14, Appl
38	20	100.0	480	4	US-09-046-604-14	Sequence 14, Appl
39	20	100.0	621	1	US-07-748-292-7	Sequence 7, Appl
40	20	100.0	622	3	US-08-867-611-34	Sequence 34, Appl
41	20	100.0	622	5	PCT-US92-06965A-4	Sequence 4, Appl
42	20	100.0	663	3	US-08-824-057-3	Sequence 3, Appl
43	20	100.0	663	4	US-09-415-582-3	Sequence 3, Appl
44	20	100.0	663	4	US-09-693-596-4	Sequence 4, Appl
45	20	100.0	738	3	US-08-867-611-35	Sequence 35, Appl

ALIGNMENTS

RESULT 1
US-08-612-973-78
; Sequence 78, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P. C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612.973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-612-973-78

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ANGSGPDQRYCWHYPPKPC 20
|||||

Db 1 ANGSGPDQRPYCHWHYPKPC 20

RESULT 2

US-08-927-597-78
; Sequence 78, Application US/08927597
; Patent No. 6245503
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-927-597-78

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ANGSGPDQRPYCHWHYPKPC 20

Db 1 ANGSGPDQRPYCHWHYPKPC 20

RESULT 3

US-08-635-886C-34
; Sequence 34, Application US/08635886C
; Patent No. 655114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555

; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-34

Query Match 100.0%; Score 20; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ANGSGPDQRPYCHWHYPKPC 20

Db 1 ANGSGPDQRPYCHWHYPKPC 20

RESULT 4

US-08-262-037-38
; Sequence 38, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hoeslin
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-262-037-38

```
Query Match      100.0%; Score 20; DB 1; Length 34;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20
   |||||
DB 1 ANGSGPDQRPYCWHPKPC 20

RESULT 5
US-08-262-037-34
; Sequence 34, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 37 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-262-037-34

Query Match      100.0%; Score 20; DB 1; Length 37;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20
   |||||
DB 9 ANGSGPDQRPYCWHPKPC 28

RESULT 6
US-08-262-037-1
; Sequence 35, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
```

```
; Sequence 10, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: No
US-08-262-037-1

Query Match      100.0%; Score 20; DB 1; Length 42;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20
   |||||
DB 9 ANGSGPDQRPYCWHPKPC 28

RESULT 7
US-08-262-037-35
; Sequence 35, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
;
US-08-262-037-35

Query Match 100.0%; Score 20; DB 1; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20
   |||||
Db 17 ANGSGPDQRPYCWHPKPC 36

RESULT 8
US-08-262-037-39
; Sequence 39, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
;
US-08-262-037-35

Query Match 100.0%; Score 20; DB 1; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20
   |||||
Db 17 ANGSGPDQRPYCWHPKPC 36

RESULT 9
US-08-262-037-36
; Sequence 36, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
;
US-08-262-037-39

Query Match 100.0%; Score 20; DB 1; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.9e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20
   |||||
Db 17 ANGSGPDQRPYCWHPKPC 36

RESULT 9
US-08-262-037-36
; Sequence 36, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
;
US-08-262-037-39
```

FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/719,819
FILING DATE: 24-June-1991
APPLICATION NUMBER: 07/667,275
FILING DATE: 11-Mar-1991
APPLICATION NUMBER: 07/651,735
FILING DATE: 07-Feb-1991
APPLICATION NUMBER: 07/558,799
FILING DATE: 26-July-1990
APPLICATION NUMBER: 07/510,153
FILING DATE: 16-April-1990
ATTORNEY/AGENT INFORMATION:
NAME: Maria C. H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/POCKET NUMBER: 1151-
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800

RESULT 12
US-08-440-103-20

```
; Sequence 20, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,103
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-103-20

Query Match 100.0%; Score 20; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 4.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQPCWHYPPKPC 20
Db 106 ANGSGPDQPCWHYPPKPC 125

RESULT 13
US-08-440-542-20
; Sequence 20, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-103-20
```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,542
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-542-20

Query Match 100.0%; Score 20; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 4.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQPCWHYPPKPC 20
Db 106 ANGSGPDQPCWHYPPKPC 125

RESULT 14
US-08-231-368-20
; Sequence 20, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-231-368-20

Query Match 100.0%; Score 20; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 4.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20
| | | | | | | | | | | | | | | | | |
Db 106 ANGSGPDQRPYCWHPKPC 125

RESULT 15

US-08-440-210-20
; Sequence 20, Application US/08440210
; Patent No. 5766845
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440.210
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231.368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-210-20

Query Match 100.0%; Score 20; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 4.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20
| | | | | | | | | | | | | | | | | |
Db 106 ANGSGPDQRPYCWHPKPC 125

Search completed: November 21, 2003, 21:15:15
Job time : 10.6 secs

THIS PAGE BLANK (ISPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time 19.5 seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-78

Perfect score: 20

Sequence: 1 ANGSGPDQRYCWHYPRKC 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	10	US-09-973-025-78
2	20	100.0	20	11	US-09-899-303-78
3	20	100.0	20	11	US-09-995-808-78
4	20	100.0	20	11	US-09-995-860-78
5	20	100.0	20	12	US-09-995-791-78
6	20	100.0	254	10	US-09-407-430-3
7	20	100.0	363	10	US-09-407-430-2
8	20	100.0	637	12	US-10-187-257-4
9	20	100.0	637	12	US-10-265-083-2
10	20	100.0	2894	10	US-09-941-611-23
11	20	100.0	2894	15	US-10-044-995-23
12	20	100.0	3011	9	US-09-916-359-2
13	20	100.0	3011	16	US-10-232-643-6
14	12	60.0	2985	15	US-10-259-275-40
15	9	45.0	29	10	US-09-921-397-24

16	9	45.0	33	10	US-09-921-397-14	Sequence 14, Appl
17	9	45.0	34	16	US-10-318-200-12	Sequence 12, Appl
18	9	45.0	37	10	US-09-921-397-36	Sequence 36, Appl
19	9	45.0	176	10	US-09-921-397-81	Sequence 81, Appl
20	9	45.0	250	10	US-09-952-572-8	Sequence 8, Appl
21	9	45.0	290	12	US-10-128-587A-3	Sequence 3, Appl
22	9	45.0	290	15	US-10-128-590-3	Sequence 5, Appl
23	9	45.0	301	12	US-10-128-587A-5	Sequence 5, Appl
24	9	45.0	301	15	US-10-128-590-5	Sequence 5, Appl
25	9	45.0	314	10	US-09-973-025-42	Sequence 42, Appl
26	9	45.0	314	11	US-09-899-303-42	Sequence 42, Appl
27	9	45.0	314	11	US-09-995-808-42	Sequence 42, Appl
28	9	45.0	314	11	US-09-995-860-42	Sequence 42, Appl
29	9	45.0	314	12	US-09-995-791-42	Sequence 42, Appl
30	9	45.0	319	10	US-09-973-025-44	Sequence 44, Appl
31	9	45.0	319	11	US-09-899-303-44	Sequence 44, Appl
32	9	45.0	319	11	US-09-995-808-44	Sequence 44, Appl
33	9	45.0	319	11	US-09-995-860-44	Sequence 44, Appl
34	9	45.0	319	12	US-09-995-791-44	Sequence 44, Appl
35	9	45.0	338	10	US-09-973-025-38	Sequence 38, Appl
36	9	45.0	338	11	US-09-899-303-38	Sequence 38, Appl
37	9	45.0	338	11	US-09-995-808-38	Sequence 38, Appl
38	9	45.0	338	11	US-09-995-860-38	Sequence 38, Appl
39	9	45.0	338	12	US-09-995-791-38	Sequence 38, Appl
40	9	45.0	343	10	US-09-973-025-40	Sequence 40, Appl
41	9	45.0	343	11	US-09-899-303-40	Sequence 40, Appl
42	9	45.0	343	11	US-09-995-808-40	Sequence 40, Appl
43	9	45.0	343	11	US-09-995-860-40	Sequence 40, Appl
44	9	45.0	343	12	US-09-995-791-40	Sequence 40, Appl
45	9	45.0	347	11	US-09-194-949-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1

US-09-973-025-78
; Sequence 78, Application US/09973025
; Publication No. US20020182706A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUUSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100

```

, INFORMATION FOR SEQ ID NO: 78:
,
, SEQUENCE CHARACTERISTICS:
,     LENGTH: 20 amino acids
,     TYPE: amino acid
,     STRANDEDNESS: single
,     TOPOLOGY: linear
,     MOLECULE TYPE: peptide
,     SEQUENCE DESCRIPTION: SEQ ID NO: 78:
US-09-973-025-78

```

```
Query Match      100.0%; Score 20; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 ANGSGDQRPYCWHPKPC 20
|||
Db 1 ANGSGDQRPYCWHPKPC 20

RESULT 2
US-09-899-303-78
; Sequence 78, Application US/09899303
; Publication No. US20030036110A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT

Query Match	100.0%;	Score 20;	DB 11;	Length 20;
Best Local Similarity	100.0%;	Pred. No. 6.8e-14;		
Matches 20:	Conservative	0;	Mismatches	0;
	Indels	0;	Gaps	0;

Qy 1 ANGSGPDQRPYCWHYPKPC 20

1 ANGSGPDORPYCWHYPKPC 20

```

RESULT 3
US-09-995-808-78
; Sequence 78, Application US/09995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patentin 3.1
; SEQ ID NO 78
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-78

```

Query Match	100.0%;	Score 20;	DB 11;	Length 20;
Best Local Similarity	100.0%;	Pred. No. 6.8e-14;		
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	

Qy 1 ANGSGPDQRPYCWHYPPKPC 20
|||||

Db 1 ANGSGPDQRPYCWHYPPKPC 20
|||||

```

RESULT 4
US-09-995-860-78
; Sequence 78, Application US/09995860
; Publication No. US20030118603A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 78
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; US-09-995-860-78

```

Query Match	100.0%;	Score 20;	DB 11;	Length 20;
Best Local Similarity	100.0%;	Pred. No. 6.8e-14;		
Matches	20;	Conservative	0;	Mismatches 0;
				Indels 0;
				Gaps 0;

Qy 1 ANGSPDQRPYCWHPKPC 20
|||||

Db 1 ANGSPDQRPYCWHPKPC 20
|||||

```

RESULT 5
US-09-995-791-78
; Sequence 78, Application US/09995791
; Publication No. US20030147918A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 78

```

```
Query Match 100.0%; Score
Best Local Similarity 100.0%; Pr
Matches 20: Conservative 0:
```

Qy 1 ANGSGPDQRPYCWHYPKPC 20

```
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-78

Query Match      100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANGSGPDORPYCWHYPPKPC 20
  |||||
Db 1 ANGSGPDORPYCWHYPPKPC 20
  |||||

RESULT 6
US-09-407-430-3
; Sequence 3, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; TITLE OF INVENTION: HEPATITIS C VIRUS INFECTION
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-407-430-3

Query Match      100.0%; Score 20; DB 10; Length 254;
Best Local Similarity 100.0%; Pred. No. 5.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANGSGPDORPYCWHYPPKPC 20
  |||||
Db 70 ANGSGPDORPYCWHYPPKPC 89
  |||||

RESULT 7
US-09-407-430-2
; Sequence 2, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; TITLE OF INVENTION: HEPATITIS C VIRUS INFECTION
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-09-407-430-2

Query Match      100.0%; Score 20; DB 10; Length 363;
Best Local Similarity 100.0%; Pred. No. 7.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANGSGPDORPYCWHYPPKPC 20
  |||||
Db 92 ANGSGPDORPYCWHYPPKPC 111
  |||||

RESULT 8
US-10-187-257-4
; Sequence 4, Application US/10187257
; Publication No. US20030138458A1
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: O'HAGAN, Derek
; TITLE OF INVENTION: HCV E1E2 VACCINE COMPOSITIONS
; FILE REFERENCE: 2302-17206
; CURRENT APPLICATION NUMBER: US/10/187,257
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
; OTHER INFORMATION: amino acid
US-10-187-257-4

Query Match      100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 1.1e-12;
Matches %0; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANGSGPDORPYCWHYPPKPC 20
  |||||
Db 303 ANGSGPDORPYCWHYPPKPC 322
  |||||

RESULT 9
US-10-265-083-2
; Sequence 2, Application US/10265083
; Publication No. US20030170273A1
; GENERAL INFORMATION:
; APPLICANT: O'HAGAN, Derek
; APPLICANT: VALIANTE, Nicholas
; TITLE OF INVENTION: ADJUVANT COMPOSITIONS
; FILE REFERENCE: 2302-18357.30
; CURRENT APPLICATION NUMBER: US/10/265,083
; CURRENT FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7
; OTHER INFORMATION: region
US-10-265-083-2

Query Match      100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 1.1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANGSGPDORPYCWHYPPKPC 20
  |||||
Db 303 ANGSGPDORPYCWHYPPKPC 322
  |||||

RESULT 10
US-09-941-611-23
; Sequence 23, Application US/09941611
; Patent No. US20020106640A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSWUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; ANTIBODIES TO HEPATITIS C VIRUS
```

NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/941,611
FILING DATE: 30-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: 1995-02-21
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-941-611-23

Query Match 100.0%; Score 20; DB 10; Length 2894;
Best Local Similarity 100.0%; Pred. No. 3.8e-12; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;

QY 1 ANSGGPDQRPYCHWYPPKPC 20
|||||
DB 475 ANSGGPDQRPYCHWYPPKPC 494

RESULT 11
US-10-044-995-23
Sequence 23, Application US/10044995
Publication No. US20030049685A1
GENERAL INFORMATION:
APPLICANT: DELEYS, ROBERT J
POLLET, DIRK
MAERTENS, GEERT
VAN HEUVERSWUN, HUGO
TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
CORRESPONDENCE ADDRESSES: 23
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/044,995
FILING DATE: 15-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: <Unknown>
APPLICATION NUMBER: US 07/920,286
FILING DATE: 14-OCT-1992
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-044-995-23

Query Match 100.0%; Score 20; DB 15; Length 2894;
Best Local Similarity 100.0%; Pred. No. 3.8e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANSGGPDQRPYCHWYPPKPC 20
|||||
DB 475 ANSGGPDQRPYCHWYPPKPC 494

RESULT 12
US-09-916-359-2
Sequence 2, Application US/09916359
Patent No. US20020034734A1
GENERAL INFORMATION:
APPLICANT: Veronique Barban
TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
TREATING C HEPATITIS
FILE REFERENCE: PFCF97-03A
CURRENT APPLICATION NUMBER: US/09/916,359
CURRENT FILING DATE: 2001-07-26
PRIOR APPLICATION NUMBER: 09/388,874
PRIOR FILING DATE: 1999-09-02
PRIOR APPLICATION NUMBER: 97/02,887
PRIOR FILING DATE: 1997-03-06
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 3011
TYPE: PRT
ORGANISM: Virus
US-09-916-359-2

Query Match 100.0%; Score 20; DB 9; Length 3011;
Best Local Similarity 100.0%; Pred. No. 4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ANSGGPDQRPYCHWYPPKPC 20
6 |||||

Db 475 ANGSGPDQPCWCHVPPKPC 494

RESULT 13
US-10-232-643-6
; Sequence 6, Application US/10232643
; Publication No. US20030129586A1
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, MICHAEL
; CHOO, QUI-LIM
; HAN, JANG
; CHOE, JOONHO
; TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
; HELICASE ACTIVITY AND IMPROVED SOLUBILITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: Intellectual Property - R440, P.O. Box 8097
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/232,643
; FILING DATE: 30-Aug-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/483,799
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/529,169
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0100.005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-3274
; TELEFAX: (510) 655-3542
; TELEX: n/a
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 9
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Lys or Arg"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 11
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Asn or Thr"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 176
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Ile or Thr"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 334
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Met or Val"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2502
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Leu or Phe"
; LOCATION: 603
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Leu or Ile"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 848
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Tyr or Asn"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1114
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Pro or Ser"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1117
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Ser or Thr"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1276
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Pro or Leu"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1454
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Cys or Tyr"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1471
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Thr or Ser"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1877
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Glu or Gly"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1948
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Leu or His"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1949
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Ser or Cys"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2021
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Gly or Val"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2349
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Thr or Ser"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2385
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Tyr or Phe"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2386
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Ser or Ala"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2502
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Leu or Phe"

```
;
;
; NAME/KEY: Duplication
; LOCATION: 2690
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Arg or Gly"
;
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2921
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Arg or Gly"
;
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2996
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Leu or Pro"
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-232-643-6
Query Match 100.0%; Score 20; DB 16; Length 3011;
Best Local Similarity 100.0%; Pred. No. 4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANGSGDQRPYCWHPKPC 20
| | | | | | | | | | | | | | | | | |
Db 475 ANGSGDQRPYCWHPKPC 494

RESULT 14
US-10-259-275-40
; Sequence 40, Application US/10259275
; Publication No. US20030125541A1
; GENERAL INFORMATION:
; APPLICANT: Lemon, Stanley M.
; APPLICANT: Yi, Minkyung
; TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE
; FILE REFERENCE: 265.0007 0120
; CURRENT APPLICATION NUMBER: US/10/259,275
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: US 60/171,909
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: US 09/747,419
; PRIOR FILING DATE: 2000-12-23
; PRIOR APPLICATION NUMBER: US 60/325,236
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/338,123
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 40
; LENGTH: 2985
; TYPE: PRT
; ORGANISM: ARTIFICIAL
; FEATURE:
; OTHER INFORMATION: amino acid sequence encoded by nucleotides 2077-11121 of SEQ ID
; OTHER INFORMATION: NO:39
US-10-259-275-40
Query Match 60.0%; Score 12; DB 15; Length 2985;
Best Local Similarity 100.0%; Pred. No. 0.00052;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 PDQRPYCWHPK 17
| | | | | | | | | | | | | | | |
Db 481 PDQRPYCWHPK 492

RESULT 15
US-09-921-397-24
; Sequence 24, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
```

```
;
;
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; applications thereof
; FILE REFERENCE: B4809A, JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-24
Query Match 45.0%; Score 9; DB 10; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 RPYCWHPK 17
| | | | | | | | | | | | | | | |
Db 2 RPYCWHPK 10

Search completed: November 21, 2003, 22:19:38
Job time : 19.5 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 163.1 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-78

Perfect score: 20

Sequence: 1 ANGSPDQPCYCHYPPKPC 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA_Main:*

- 1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/1/paa/US097A_COMB.pep.*
- 22: /cgn2_6/ptodata/1/paa/US097B_COMB.pep.*
- 23: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
- 24: /cgn2_6/ptodata/1/paa/US099A_COMB.pep.*
- 25: /cgn2_6/ptodata/1/paa/US099B_COMB.pep.*
- 26: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
- 27: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
- 28: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
- 29: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
- 30: /cgn2_6/ptodata/1/paa/US104_COMB.pep.*
- 31: /cgn2_6/ptodata/1/paa/US106_COMB.pep.*
- 32: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	13	US-08-974-685-34
					Sequence 34, Appl

2	20	100.0	20	13	US-08-974-690-34	Sequence 34, Appl
3	20	100.0	20	13	US-08-974-690A-34	Sequence 34, Appl
4	20	100.0	20	13	US-08-974-690B-34	Sequence 34, Appl
5	20	100.0	20	13	US-08-974-690C-34	Sequence 34, Appl
6	20	100.0	20	23	US-09-899-303-78	Sequence 78, Appl
7	20	100.0	20	23	US-09-899-303A-78	Sequence 78, Appl
8	20	100.0	20	25	US-09-973-025-78	Sequence 78, Appl
9	20	100.0	20	25	US-09-995-791-78	Sequence 78, Appl
10	20	100.0	20	25	US-09-995-808-78	Sequence 78, Appl
11	20	100.0	20	25	US-09-995-860-78	Sequence 78, Appl
12	20	100.0	20	26	US-10-020-510-78	Sequence 78, Appl
13	20	100.0	20	29	US-10-321-798-78	Sequence 78, Appl
14	20	100.0	34	8	US-08-475-482-38	Sequence 38, Appl
15	20	100.0	34	8	US-08-477-072-38	Sequence 38, Appl
16	20	100.0	34	8	US-08-477-582-38	Sequence 38, Appl
17	20	100.0	34	8	US-08-480-253-38	Sequence 34, Appl
18	20	100.0	37	8	US-08-475-482-34	Sequence 34, Appl
19	20	100.0	37	8	US-08-477-072-34	Sequence 34, Appl
20	20	100.0	37	8	US-08-477-582-34	Sequence 34, Appl
21	20	100.0	37	8	US-08-480-253-34	Sequence 34, Appl
22	20	100.0	42	3	US-07-667-275A-1	Sequence 1, Appl
23	20	100.0	42	8	US-08-475-482-1	Sequence 1, Appl
24	20	100.0	42	8	US-08-477-072-1	Sequence 1, Appl
25	20	100.0	42	8	US-08-477-582-1	Sequence 1, Appl
26	20	100.0	42	8	US-08-480-253-1	Sequence 35, Appl
27	20	100.0	45	8	US-08-475-482-35	Sequence 35, Appl
28	20	100.0	45	8	US-08-477-072-35	Sequence 35, Appl
29	20	100.0	45	8	US-08-477-582-35	Sequence 35, Appl
30	20	100.0	45	8	US-08-480-253-35	Sequence 35, Appl
31	20	100.0	50	8	US-08-475-482-39	Sequence 39, Appl
32	20	100.0	50	8	US-08-477-072-39	Sequence 39, Appl
33	20	100.0	50	8	US-08-477-582-39	Sequence 39, Appl
34	20	100.0	50	8	US-08-480-253-39	Sequence 39, Appl
35	20	100.0	52	8	US-08-475-482-36	Sequence 36, Appl
36	20	100.0	52	8	US-08-477-072-36	Sequence 36, Appl
37	20	100.0	52	8	US-08-477-582-36	Sequence 36, Appl
38	20	100.0	52	8	US-08-480-253-36	Sequence 40, Appl
39	20	100.0	57	8	US-08-475-482-40	Sequence 40, Appl
40	20	100.0	57	8	US-08-477-072-40	Sequence 40, Appl
41	20	100.0	57	8	US-08-477-582-40	Sequence 40, Appl
42	20	100.0	57	8	US-08-480-253-40	Sequence 40, Appl
43	20	100.0	76	8	US-08-403-590B-79	Sequence 79, Appl
44	20	100.0	76	8	US-08-444-112-79	Sequence 79, Appl
45	20	100.0	122	8	US-08-436-966-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1

US-08-974-685-34

Sequence 34, Application US/08974685

GENERAL INFORMATION:

APPLICANT: LEROUX-ROELS, GEERT

DELEYS, ROBERT

MAERTENS, GEERT

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF

HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 181

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHVE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,685
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-08-974-685-34

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANGSGPDQPCYCHYPPKPC 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 ANGSGPDQPCYCHYPPKPC 20

RESULT 2

US-08-974-690-34
Sequence 34, Application US/08974690
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
APPLICANT: MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 166
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/635,886
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

US-08-974-690-34

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANGSGPDQPCYCHYPPKPC 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 ANGSGPDQPCYCHYPPKPC 20

RESULT 3

US-08-974-690A-34
Sequence 34, Application US/08974690A
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
APPLICANT: MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690A
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-08-974-690A-34

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANGSGPDQPCYCHYPPKPC 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 ANGSGPDQPCYCHYPPKPC 20

RESULT 4

US-08-974-690B-34
Sequence 34, Application US/08974690B
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
APPLICANT: MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303A
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 78:
US-09-899-303A-78
Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ANGSGPDQRPYCWHPKPC 20
Db 1 ANGSGPDQRPYCWHPKPC 20
RESULT 8
US-09-973-025-78
Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ANGSGPDQRPYCWHPKPC 20
Db 1 ANGSGPDQRPYCWHPKPC 20
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/973,025
FILING DATE: 10-Oct-2001
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 78:
US-09-973-025-78
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ANGSGPDQRPYCWHPKPC 20
Db 1 ANGSGPDQRPYCWHPKPC 20
RESULT 9
US-09-995-791-78
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ANGSGPDQRPYCWHPKPC 20
Db 1 ANGSGPDQRPYCWHPKPC 20
GENERAL INFORMATION:
APPLICANT: Innogenetics N.V.
TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
TITLE OF INVENTION: therapeutic use.
FILE REFERENCE: 2551-68
CURRENT APPLICATION NUMBER: US/09/995,791
CURRENT FILING DATE: 2001-11-29
NUMBER OF SEQ ID NOS: 122
SOFTWARE: Patent In 3.1
SEQ ID NO 78
LENGTH: 20
TYPE: PRT
ORGANISM: Hepatitis C virus
US-09-995-791-78
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ANGSGPDQRPYCWHPKPC 20
Db 1 ANGSGPDQRPYCWHPKPC 20
RESULT 10
US-09-995-808-78
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ANGSGPDQRPYCWHPKPC 20
Db 1 ANGSGPDQRPYCWHPKPC 20
GENERAL INFORMATION:
APPLICANT: Innogenetics N.V.
TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
TITLE OF INVENTION: therapeutic use.
FILE REFERENCE: 2551-70
CURRENT APPLICATION NUMBER: US/09/995,808
CURRENT FILING DATE: 2001-11-29
NUMBER OF SEQ ID NOS: 122
SOFTWARE: Patent In 3.1
SEQ ID NO 78
LENGTH: 20
TYPE: PRT
ORGANISM: Hepatitis C virus
US-09-995-808-78
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;

```
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ANGSGDQRPYCWHPKPC 20
    |||||
DB 1 ANGSGDQRPYCWHPKPC 20
    |||||

RESULT 11
US-09-995-860-78
; Sequence 78, Application US/0995860
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 78
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-78

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ANGSGDQRPYCWHPKPC 20
    |||||
DB 1 ANGSGDQRPYCWHPKPC 20
    |||||

RESULT 12
US-10-020-510-78
; Sequence 78, Application US/10020510
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-72
; CURRENT FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 78
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-020-510-78

Query Match 100.0%; Score 20; DB 26; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ANGSGDQRPYCWHPKPC 20
    |||||
DB 1 ANGSGDQRPYCWHPKPC 20
    |||||

RESULT 13
US-10-321-798-78
; Sequence 78, Application US/10321798
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-93
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: 60/418,358
; PRIOR FILING DATE: 2002-10-16
```

```
; PRIOR APPLICATION NUMBER: 10/020,510
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 78
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-321-798-78

Query Match 100.0%; Score 20; DB 29; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ANGSGDQRPYCWHPKPC 20
    |||||
DB 1 ANGSGDQRPYCWHPKPC 20
    |||||

RESULT 14
US-08-475-482-38
; Sequence 38, Application US/08475482
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & PINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION NUMBER: US/08/475,482
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 17-June-1994
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-475-482-38
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.55 Seconds
(without alignments)
117.536 Million cell updates/sec

Title: US-09-973-025-78

Perfect score: 20

Sequence: 1 ANGSGPDQRPYCHWYPPKPC 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA_New.*

- 1: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US11_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US12_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	6	US-10-651-165-34
2	20	100.0	333	1	PCT-US03-19834-3
3	20	100.0	333	6	US-10-655-562-4
4	20	100.0	637	1	PCT-US03-33610-4
5	20	100.0	3011	1	PCT-US03-19834-2
6	20	100.0	3011	6	US-10-296-734-406
7	19	95.0	30	30	US-10-296-734-470
8	19	95.0	5985	6	US-10-296-734-814
9	19	95.0	5985	6	US-10-296-734-810
10	16	80.0	30	6	US-10-296-734-472
11	11	55.0	1026	1	PCT-US03-20409-3
12	9	45.0	34	6	US-10-685-435-21
13	9	45.0	347	6	US-10-664-391-9
14	9	45.0	539	6	US-10-664-391-11
15	9	45.0	2280	1	PCT-US03-20322-211
16	9	45.0	3033	6	US-10-009-002-5
17	8	40.0	20	6	US-10-651-165-33
18	8	40.0	20	6	US-10-651-165-35
19	8	40.0	176	5	US-09-638-693-158
20	8	40.0	176	5	US-09-873-224A-158
21	7	35.0	28	6	US-10-685-435-20
22	6	30.0	14	5	US-09-638-693-267
23	6	30.0	14	5	US-09-873-224A-267
24	6	30.0	136	1	PCT-US03-23932-4
25	6	30.0	238	6	US-10-679-063-17338
26	6	30.0	322	6	US-10-425-114A-44496

27	6	30.0	330	6	US-10-425-114A-43742	Sequence 43742, A
28	6	30.0	495	6	US-10-425-114A-50093	Sequence 50093, A
29	6	30.0	591	6	US-10-679-063-894	Sequence 894, App
30	6	30.0	604	6	US-10-679-063-7007	Sequence 7007, App
31	6	30.0	605	6	US-10-679-063-22608	Sequence 22608, A
32	6	30.0	607	6	US-10-679-063-14536	Sequence 14536, A
33	6	30.0	612	6	US-10-679-063-25564	Sequence 25564, A
34	6	30.0	659	6	US-10-425-114A-55579	Sequence 55579, A
35	5	25.0	30	6	US-10-685-435-41	Sequence 41, Appl
36	5	25.0	57	6	US-10-425-114A-45150	Sequence 45150, A
37	5	25.0	80	6	US-10-679-063-5015	Sequence 5015, App
38	5	25.0	96	6	US-10-425-114A-56270	Sequence 56270, A
39	5	25.0	97	6	US-10-425-114A-56726	Sequence 56726, A
40	5	25.0	97	6	US-10-425-114A-69207	Sequence 69207, A
41	5	25.0	98	6	US-10-425-114A-45009	Sequence 45009, A
42	5	25.0	100	6	US-10-679-063-4488	Sequence 4488, App
43	5	25.0	105	6	US-10-425-114A-61830	Sequence 61830, A
44	5	25.0	106	6	US-10-425-114A-63148	Sequence 63148, A
45	5	25.0	110	6	US-10-425-114A-41508	Sequence 41508, A

ALIGNMENTS

RESULT 1

US-10-651-165-34
; Sequence 34, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
; LENGTH: 20
; TYPE: PRT

US-10-651-165-34

Query Match 100.0%; Score 20; DB 6; Length 20;

Best Local Similarity 100.0%; Pred. No. 7.2e-10; Mismatches 0; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCHWYPPKPC 20

DB 1 ANGSGPDQRPYCHWYPPKPC 20

RESULT 2

PCT-US03-19834-3

; Sequence 3, Application PC/TUS0319834

; GENERAL INFORMATION:

; APPLICANT: STAPLETON, JACK T.

; APPLICANT: WUENSCHMANN, SABINA

; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND

; TITLE OF INVENTION: METHODS FOR TREATING HYPERLIPIDEMIA

; FILE REFERENCE: IOWA:045WO

; CURRENT APPLICATION NUMBER: PCT/US03/19834

; CURRENT FILING DATE: 2003-06-24

; PRIOR APPLICATION NUMBER: 10/445,724

; PRIOR FILING DATE: 2003-05-27

; PRIOR APPLICATION NUMBER: 60/392,158

```
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-3

Query Match      100.0%; Score 20; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 2.4e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 ANGSGDQRPYCWHPKPC 20
Db      93 ANGSGDQRPYCWHPKPC 112

RESULT 3
US-10-655-562-4
; Sequence 4, Application US/10655562
; GENERAL INFORMATION:
; APPLICANT: BISHOP-HURLEY, SHARON L.
; APPLICANT: SCHMIDT, FRANCIS J.
; APPLICANT: SMITH, ARNOLD L.
; TITLE OF INVENTION: PHAGE-DISPLAY PEPTIDES AS NOVEL ANTIMICROBIAL AGENTS
; FILE REFERENCE: UOVO:022US
; CURRENT APPLICATION NUMBER: US/10/655,562
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: 60,409,909
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-10-655-562-4

Query Match      100.0%; Score 20; DB 6; Length 333;
Best Local Similarity 100.0%; Pred. No. 2.4e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 ANGSGDQRPYCWHPKPC 20
Db      93 ANGSGDQRPYCWHPKPC 112

RESULT 4
PCT-US03-33610-4
; Sequence 4, Application PC/TUS0333610
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: SELBY, Mark
; APPLICANT: PALIARD, Xavier
; TITLE OF INVENTION: ACTIVATION OF HCV-SPECIFIC T-CELLS
; FILE REFERENCE: 2300-1612.60
; CURRENT APPLICATION NUMBER: PCT/US03/33610
; CURRENT FILING DATE: 2003-10-24
; PRIOR APPLICATION NUMBER: 10/281,341
; PRIOR FILING DATE: 2002-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
PCT-US03-33610-4

Query Match      100.0%; Score 20; DB 1; Length 637;
Best Local Similarity 100.0%; Pred. No. 3.2e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 ANGSGDQRPYCWHPKPC 20
Db      303 ANGSGDQRPYCWHPKPC 322

RESULT 5
PCT-US03-19834-2
; Sequence 2, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; FILE REFERENCE: IOWA:045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-2

Query Match      100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 6.2e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 ANGSGDQRPYCWHPKPC 20
Db      475 ANGSGDQRPYCWHPKPC 494

RESULT 6
US-10-296-734-406
; Sequence 406, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296.734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a consensus polypeptide
US-10-296-734-406

Query Match      100.0%; Score 20; DB 6; Length 3011;
Best Local Similarity 100.0%; Pred. No. 6.2e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 ANGSGDQRPYCWHPKPC 20
Db      475 ANGSGDQRPYCWHPKPC 494
```

```

; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 810
; LENGTH: 5985
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a savine
US-10-296-734-810

Query Match          95.0%; Score 19; DB 6; Length 5985;
Best Local Similarity 100.0%; Pred. No. 5e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGDQRPYCHWYPPKP 19
   |||||
Db 3972 ANGSGDQRPYCHWYPPKP 3990
   |||||

RESULT 10
US-10-296-734-472
; Sequence 472, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2003-08-04
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 472
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 33
US-10-296-734-472

Query Match          80.0%; Score 16; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GPDQRPYCHWYPPKP 20
   |||||
Db 1 GPDQRPYCHWYPPKP 16
   |||||

RESULT 11
PCT-US03-20409-3
; Sequence 3, Application PC/TUS0320409
; GENERAL INFORMATION:
; APPLICANT: Merck & Co., Inc.
; TITLE OF INVENTION: HEPATITIS C VIRUS PARTICLE FORMATION
; FILE REFERENCE: 21123 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/20409
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/393,167
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1026
; TYPE: PRT
; ORGANISM: HCV
PCT-US03-20409-3

Query Match          55.0%; Score 11; DB 1; Length 1026;

```

Best Local Similarity 100.0%; Pred. No. 0.037;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 DORPYCWHYPP 17
Db 481 DORPYCWHYPP 491

RESULT 12
US-10-685-435-21
; Sequence 21, Application US/10685435
; GENERAL INFORMATION:
; APPLICANT: Immunogenetics N.V.
; TITLE OF INVENTION: Multi-mer peptides derived from Hepatitis C Virus
; TITLE OF INVENTION: envelope proteins for diagnostic use and vaccination
; TITLE OF INVENTION: purposes.
; FILE REFERENCE: PCT98.75.HCV30
; CURRENT APPLICATION NUMBER: US/10/685,435
; CURRENT FILING DATE: 2003-10-16
; PRIOR APPLICATION NUMBER: US/09/566,266B
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 97870179.5
; PRIOR FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: patenting Ver. 2.1
; SEQ ID NO 21
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Hepatitis C virus

US-10-685-435-21
Query Match 45.0%; Score 9; DB 6; Length 34;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 DORPYCWHY 15
Db 2 DORPYCWHY 10

RESULT 13
US-10-664-391-9
; Sequence 9, Application US/10664391
; GENERAL INFORMATION:
; APPLICANT: Donnelly, John J.
; APPLICANT: Liu, Margaret A.
; APPLICANT: Shiver, John W.
; APPLICANT: Fu, Tong-Ming
; TITLE OF INVENTION: SYNTHETIC HEPATITIS C GENES
; FILE REFERENCE: 19732YPCA
; CURRENT APPLICATION NUMBER: US/10/664,391
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: PCT/US97/09884
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/033,534
; PRIOR FILING DATE: 1996-12-20
; PRIOR APPLICATION NUMBER: 60/020,494
; PRIOR FILING DATE: 1996-06-11
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Hepatitis C Virus

US-10-664-391-9
Query Match 45.0%; Score 9; DB 6; Length 347;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 DORPYCWHY 15
Db 99 DORPYCWHY 107

RESULT 14
US-10-664-391-11
; Sequence 11, Application US/10664391
; GENERAL INFORMATION:
; APPLICANT: Donnelly, John J.
; APPLICANT: Liu, Margaret A.
; APPLICANT: Shiver, John W.
; APPLICANT: Fu, Tong-Ming
; TITLE OF INVENTION: SYNTHETIC HEPATITIS C GENES
; FILE REFERENCE: 19732YPCA
; CURRENT APPLICATION NUMBER: US/10/664,391
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: PCT/US97/09884
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/033,534
; PRIOR FILING DATE: 1996-12-20
; PRIOR APPLICATION NUMBER: 60/020,494
; PRIOR FILING DATE: 1996-06-11
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Hepatitis C Virus

US-10-664-391-11
Query Match 45.0%; Score 9; DB 6; Length 539;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 DORPYCWHY 15
Db 291 DORPYCWHY 299

RESULT 15
PCT-US03-20322-211
; Sequence 211, Application PC/TUS0320322
; GENERAL INFORMATION:
; APPLICANT: Dana-Farber Cancer Institute, Inc., et al.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
; FILE REFERENCE: DEN-043PC
; CURRENT APPLICATION NUMBER: PCT/US03/20322
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 211
; LENGTH: 2280
; TYPE: PRT
; ORGANISM: Hepatitis C Virus

PCT-US03-20322-211
Query Match 45.0%; Score 9; DB 1; Length 2280;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 DORPYCWHY 15
Db 481 DORPYCWHY 489

Search completed: November 21, 2003, 22:12:55
Job time : 9.55 secs

Db 475 ANGSGDQRPYCWHPKPC 494

RESULT 2

S40770

genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: S40770; PC1285
R:Okamoto, H.
submitted to the EMBL Data Library, March 1992
A:Reference number: S40770
A:Accession: S40770
A:Molecule type: genomic RNA
A:Residues: 1-3011 <OKA>
A:Cross-references: EMBL:D10749; NID:G221586; PIDN:BAA01582.1; PID:G221587
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; MUID:91013116; PMID:2170712
A:Accession: PC1285
A:Molecule type: genomic RNA
A:Residues: 1-513 <OK2>
A:Cross-references: GB:D00831; NID:G221511; PIDN:BAA00705.1; PID:G221512
A:Experimental source: isolate HC-J1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1662/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGDQRPYCWHPKPC 20
|||||

Db 475 ANGSGDQRPYCWHPKPC 494

RESULT 3

A44150

structural protein - hepatitis C virus
C:Species: hepatitis C virus
C:Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 17-Nov-2000
C:Accession: A44150
R:Ching, W.M.; Wychowski, C.; Beach, M.J.; Wang, H.; Davies, C.L.; Carl, M.; Bradley, D. Proc. Natl. Acad. Sci. U.S.A. 89, 3190-3194, 1992
A:Title: Interaction of immune sera with synthetic peptides corresponding to the structural protein of hepatitis C virus.
A:Reference number: A44150; MUID:92228749; PMID:1373489
A:Accession: A44150
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: genomic RNA
A:Residues: 1-513 <CHI>
C:Superfamily: hepatitis C virus genome polyprotein

Query Match 95.0%; Score 19; DB 2; Length 513;
Best Local Similarity 100.0%; Pred. No. 3.2e-13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGDQRPYCWHPKPC 19
|||||

Db 475 ANGSGDQRPYCWHPKPC 493

RESULT 4

S24080

envelope protein - hepatitis C virus (fragment)
N:Contains: envelope protein gp35; envelope protein gp70
C:Species: hepatitis C virus
C:Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C:Accession: S24080
R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraishi, K.; Ohkoshi, Virus Res. 22, 107-123, 1992
A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A:Reference number: S24068; MUID:92230394; PMID:1314471
A:Accession: S24080
A:Molecule type: genomic RNA
A:Residues: 1-138 <KAT>
A:Cross-references: EMBL:X60568
A:Experimental source: isolate RE42
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; glycoprotein; polyprotein
F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted
F:49,55,62,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 80.0%; Score 16; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 1.9e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GPDQRPYCWHPKPC 20
|||||

Db 111 GPDQRPYCWHPKPC 126

RESULT 5

S24081

envelope protein - hepatitis C virus (fragment)
N:Contains: envelope protein gp35; envelope protein gp70
C:Species: hepatitis C virus
C:Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C:Accession: S24081
R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraishi, K.; Ohkoshi, Virus Res. 22, 107-123, 1992
A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A:Reference number: S24068; MUID:92230394; PMID:1314471
A:Accession: S24081
A:Molecule type: genomic RNA
A:Residues: 1-138 <KAT>
A:Cross-references: EMBL:X60569
A:Experimental source: isolate RE43
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; glycoprotein; polyprotein
F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted
F:49,55,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 75.0%; Score 15; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 2.3e-09;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GSGDQRPYCWHPKPC 17
|||||

Db 109 GSGDQRPYCWHPKPC 123

RESULT 6

S24075

envelope protein - hepatitis C virus (fragment)
N:Contains: envelope protein gp35; envelope protein gp70

C:Species: hepatitis C virus
C:Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C:Accession: S24075
R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraisho, K.; Ohkoshi, Virus Res. 22, 107-123, 1992
A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A:Reference number: S24068; MUID:92230394; PMID:1314471
A:Accession: S24075
A:Molecule type: genomic RNA
A:Residues: 1-138 <KAT>
A:Experimental source: EMBL:X60564
A:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; glycoprotein; polyprotein
F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted
F:49,55,62,80,108/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 60.0%; Score 12; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 4.1e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 RPYCWHYPPKPC 20
|||||
Db 115 RPYCWHYPPKPC 126

RESULT 7
S24074
envelope protein - hepatitis C virus (fragment)
N:Contains: envelope protein gp35; envelope protein gp70
C:Species: hepatitis C virus
C:Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C:Accession: S24074
R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraisho, K.; Ohkoshi, Virus Res. 22, 107-123, 1992
A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A:Reference number: S24068; MUID:92230394; PMID:1314471
A:Accession: S24074
A:Molecule type: genomic RNA
A:Residues: 1-138 <KAT>
A:Experimental source: EMBL:X60563
A:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; glycoprotein; polyprotein
F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted
F:49,55,62,80,108/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 60.0%; Score 12; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 4.1e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 RPYCWHYPPKPC 20
|||||
Db 115 RPYCWHYPPKPC 126

RESULT 8
JQ1584
genome polyprotein - hepatitis C virus (strain U.K.) (fragment)
N:Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural protein
C:Species: hepatitis C virus
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
C:Accession: JQ1584
R:Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.
J. Gen. Virol. 73, 1521-1525, 1992
A:Title: Cloning and sequencing of the structural region and expression of putative core
A:Reference number: JQ1584; MUID:92300349; PMID:1318944

A:Accession: JQ1584
A:Molecule type: genomic RNA
A:Residues: 1-640 <KUM>
A:Cross-references: GB:X84079; NID:G643119; PIDN:CAA58888.1; PID:G643120
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; polyprotein
F:1-191/Product: core protein C #status predicted <CPC>
F:192-389/Product: envelope protein E1 #status predicted <EE1>
F:390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predicted <NS1>
F:196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn) (covalent) #status predicted <NS5>

Query Match 60.0%; Score 12; DB 2; Length 640;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 RPYCWHYPPKPC 20
|||||
Db 483 RPYCWHYPPKPC 494

RESULT 9
JQ1366
polyprotein - hepatitis C virus (French isolate) (fragments)
C:Species: hepatitis C virus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C:Accession: JQ1366
R:Kremsdorf, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.
J. Gen. Virol. 72, 2557-2561, 1991
A:Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication
A:Reference number: JQ1366; MUID:92013977; PMID:1655961
A:Accession: JQ1366
A:Molecule type: genomic RNA
A:Residues: 1-716 <KRE>
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: glycoprotein; polyprotein
F:84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #status predicted <NS5>

Query Match 60.0%; Score 12; DB 2; Length 716;
Best Local Similarity 100.0%; Pred. No. 1.6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 RPYCWHYPPKPC 20
|||||
Db 150 RPYCWHYPPKPC 161

RESULT 10
JQ0883
genome polyprotein - hepatitis C virus (strain J7) (fragments)
N:Contains: NS5 protein
C:Species: hepatitis C virus
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 23-Mar-2001
C:Accession: JQ0883
R:Okamoto, H.
Submitted to JIPID, January 1991
A:Description: The 5'-terminal and 3'-terminal sequences of the genomic RNA of hepatitis C virus
A:Reference number: JQ0879
A:Accession: JQ0883
A:Molecule type: genomic RNA
A:Residues: 1-874 <OKA>
A:Experimental source: strain J7
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein
F:510-874/Product: NS5 protein (fragment) #status predicted <NS5>

Query Match 60.0%; Score 12; DB 2; Length 874;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 RPYCWHYPPKPC 20
|||||
Db 485 RPYCWHYPPKPC 496

RESULT 11
S24100
envelope protein - hepatitis C virus (fragment)
N:Contains: envelope protein gp35; envelope protein gp70
C:Species: hepatitis C virus
C:Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C:Accession: S24100; S78234
R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraio, K.; Ohkoshi, Y. Virol. 22, 107-123, 1992
A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A:Reference number: S24068; MUID:92230394; PMID:1314471
A:Accession: S24100
A:Molecule type: genomic RNA
A:Residues: 1-138 <KAT>
A:Cross-references: EMBL:X60583
A:Experimental source: isolate RE71
R:Kato, N.
submitted to the EMBL Data Library, August 1991
A:Reference number: S78226
A:Accession: S78234
A:Molecule type: genomic RNA
A:Residues: 1-83, 'C', 85-138 <KAW>
A:Cross-references: EMBL:X60583
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; glycoprotein; polyprotein
F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted
F:49,55,62,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.0%; Score 11; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 4.9e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DORPCWHYPP 17
|||||
DB 113 DORPCWHYPP 123

RESULT 12
S24076
envelope protein - hepatitis C virus (fragment)
N:Contains: envelope protein gp35; envelope protein gp70
C:Species: hepatitis C virus
C:Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C:Accession: S24076
R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraio, K.; Ohkoshi, Y. Virol. 22, 107-123, 1992
A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A:Reference number: S24068; MUID:92230394; PMID:1314471
A:Accession: S24076
A:Molecule type: genomic RNA
A:Residues: 1-138 <KAT>
A:Cross-references: EMBL:X60556
A:Experimental source: isolate RE38
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; glycoprotein; polyprotein
F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted
F:49,55,62,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.0%; Score 11; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 4.9e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GPDQPCWHY 15
|||||
DB 111 GPDQPCWHY 121

RESULT 13

GNVWTC

genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
C:Accession: A38465
R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J. Virol. 65, 1105-1113, 1991
A:Title: Structure and organization of the hepatitis C virus genome isolated from human
A:Reference number: A38465; MUID:91140698; PMID:1847440
A:Accession: A38465
A:Molecule type: genomic RNA
A:Residues: 1-3010 <TAK>
A:Cross-references: EMBL:M58335; MID:g329770; PID:AAAY2945.1; PID:g329771
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein; capsid protein C #status predicted <CPC>
F:2-115/Product: capsid protein M #status predicted <EPM>
F:116-191/Product: envelope protein E #status predicted <MEE>
F:192-389/Product: major envelope protein NS1 #status predicted <NS1>
F:390-729/Product: nonstructural protein NS2 #status predicted <NS2>
F:730-1006/Product: nonstructural protein NS3 #status predicted <NS3>
F:1007-1615/Product: hepatitis C virus genome polyprotein A (p-loop)
F:1230-1237/Region: nucleotide-binding motif A (p-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,22

Query Match 55.0%; Score 11; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 0.00065;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DORPCWHYPP 17
|||||
DB 481 DORPCWHYPP 491

RESULT 14

C48776

polyprotein (E2/NS1 region, HVR1, HVR2) - hepatitis C virus (fragment)
C:Species: hepatitis C virus
C:Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 17-Nov-2000
C:Accession: C48776
R:Higashi, Y.; Kakum, S.; Yoshioka, K.; Wakita, T.; Mizokami, M.; Ohba, K.; Ito, Y.; Is Virol. 197, 659-668, 1993
A:Title: Dynamics of genome change in the E2/NS1 region of hepatitis C virus in vivo.
A:Reference number: A48776; MUID:94069940; PMID:8249288
A:Accession: C48776
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-133 <HIG>
A:Experimental source: subtype III, patient TR
A:Note: sequence extracted from NCBI backbone (NCBI:140216)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein

Query Match 50.0%; Score 10; DB 2; Length 133;
Best Local Similarity 100.0%; Pred. No. 0.00058;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 RPYCWHYPPK 18
|||||
DB 114 RPYCWHYPPK 123

RESULT 15

D48776

polyprotein (E2/NS1 region, HVR1, HVR2) - hepatitis C virus (fragment)

C;Species: hepatitis C virus
C;Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 17-Nov-2000
C;Accession: D48776
R;Higashi, Y.; Kakumu, S.; Yoshioka, K.; Wakita, T.; Mizokami, M.; Ohba, K.; Ito, Y.; Ise
Virology 197, 659-668, 1993
A;Title: Dynamics of genome change in the E2/NS1 region of hepatitis C virus in vivo.
A;Reference number: A48776; MUID:94069940; PMID:8249288
A;Accession: D48776
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-133 <HIG>
A;Experimental source: subtype III, patient KS
A;Note: sequence extracted from NCBI backbone (NCBIP:140217)
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: polyprotein

Query Match 50.0%; Score 10; DB 2; Length 133;
Best Local Similarity 100.0%; Pred. No. 0.00058;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 9 RPYCWHYPPK 18
| | | | | | | | | |
Db 114 RPYCWHYPPK 123

Search completed: November 21, 2003, 21:11:31
Job time : 9.25 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:52 ; Search time 4.85 Seconds
(without alignments)
193.925 Million cell updates/sec

Title: US-09-973-025-78

Perfect score: 20

Sequence: 1 ANGSGDQRPCYCWYPPKPC 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	3011	1 POLG_HCV1	P26664 h genome po
2	12	60.0	737	1 POLG_HCVJ5	P27960 hepatitis c
3	12	60.0	737	1 POLG_HCVJ7	P27961 hepatitis c
4	11	55.0	3010	1 POLG_HCVBK	P26663 h genome po
5	9	45.0	3010	1 POLG_HCVJA	P26662 h genome po
6	9	45.0	3010	1 POLG_HCVJT	Q00269 h genome po
7	9	45.0	3010	1 POLG_HCVTW	P29846 h genome po
8	9	45.0	3011	1 POLG_HCVH	P27958 h genome po
9	9	45.0	3033	1 POLG_HCVJ6	P26660 h genome po
10	9	45.0	3033	1 POLG_HCVJ8	P26661 h genome po
11	8	40.0	520	1 POLG_HCVH4	Q01404 hepatitis c
12	8	40.0	520	1 POLG_HCVHK	Q01403 hepatitis c
13	6	30.0	62	1 PLR_RANPI	Q90097 rana pipien
14	6	30.0	424	1 EVX2_HETFR	Q91a18 heterodontu
15	6	30.0	513	1 POLG_HCVJ2	P27959 hepatitis c
16	6	30.0	539	1 STG2_HUMAN	Q9ny99 homo sapien
17	6	30.0	601	1 CIK5_MUSPF	P79197 mustela put
18	6	30.0	879	1 ORP5_HUMAN	Q9h0x9 homo sapien
19	5	25.0	36	1 PYY_AMICA	P29205 amia calva
20	5	25.0	36	1 PYY_LEPSP	P09473 lepisosteus
21	5	25.0	36	1 PYY_ONCKI	P09474 oncorhynch
22	5	25.0	36	1 PYY_ORENI	P81028 oreochromis
23	5	25.0	36	1 PYY_RAJRH	P29206 raja rhina
24	5	25.0	36	1 PYY_RANRI	P29204 rana ridibu
25	5	25.0	36	1 SPY2_PHYBI	P80952 phyllomedus
26	5	25.0	37	1 PYY_CHICK	P29203 gallus gall
27	5	25.0	69	1 PYY_LOFAM	P09475 lophius ame
28	5	25.0	81	1 RELX_DASSA	P81191 dasyatis sa
29	5	25.0	97	1 PYY_BRARE	Q918p2 brachydanio
30	5	25.0	97	1 PY_DICLA	Q9pt98 dicentrarch
31	5	25.0	143	1 MCS_MOUSE	P15265 mus musculu
32	5	25.0	145	1 MCS_RAT	Q64298 rattus norv
33	5	25.0	162	1 ISPF_YERPE	Q8zbp7 yersinia pe

RESULT 1

ID	POLG_HCV1	STANDARD;	PRT;	3011 AA.
AC	P26664;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)]			
OS	Hepatitis C virus (isolate 1) (HCV)			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.			
OX	NCBI_TaxID=11104;			
RN	[1]_TaxID=11104;			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91172826; PubMed=1848704;			
RA	Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;			
RT	"Genetic organization and diversity of the hepatitis C virus.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).			
CC	-!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.			
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	-!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA)(N).			
CC	-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.			
CC	-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; M62321; AAA45676.1; -			
DR	PIR; A39166; GNMVVC3.			
DR	PDB; 1A1V; 16-FEB-99.			
DR	PDB; 1HEI; 25-NOV-98.			
DR	MEROPS; S29.001; -			
DR	MEROPS; U39.001; -			
DR	InterPro; IPR001410; DEAD.			
DR	InterPro; IPR002522; HCV_capsid.			

ALIGNMENTS

34	5	25.0	202	1	UMPK_ARATH	O04905 arabidopsis
35	5	25.0	206	1	SOD2_PLEBO	P50059 plectonema
36	5	25.0	211	1	WIN2_SOLTU	P09762 solanum tub
37	5	25.0	225	1	SRD1_YEAST	P09007 saccharomyc
38	5	25.0	252	1	NUOE_MYCTU	P95177 mycobacteri
39	5	25.0	258	1	SLA2_MOUSE	Q8410 mus musculu
40	5	25.0	259	1	ROB1_AGRRH	P20402 agrobacteri
41	5	25.0	277	1	Y31K_SSV1	P20203 sulfolobus
42	5	25.0	279	1	TNF6_MOUSE	P41047 mus musculu
43	5	25.0	280	1	DHAA_MYCSM	Q938b4 mycobacteri
44	5	25.0	280	1	TRUA_TREPA	O83802 treponema p
45	5	25.0	297	1	PTR_METFE	Q49168 methanother

DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV NS5b.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR007095; RNA pol DS P5.
 DR InterPro; IPR007094; RNA pol psivir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; Helicase C; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 MATRIX PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 224 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 5.8e-14;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ANGSGPDQRPYCVHYPKPC 20

Db 475 ANGSGPDQRPYCVHYPKPC 494
 RESULT 2
 POLG HCVJ5 STANDARD; PRT; 737 AA.
 AC P27950;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
 DE protein [Envelope protein M]; Major envelope protein E; Nonstructural
 DE proteins NS1 and NS2] (Fragment).
 OS Hepatitis C virus (isolate HC-J5) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11112;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kural K., Okada S.-I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MNNA.
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D10075; BAA00969.1; -;
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 0 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 MATRIX PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 733 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 734 >737 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).


```

FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 737 737
SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CRC64;

Query Match 60.0%; Score 12; DB 1; Length 737;
Best Local Similarity 100.0%; Pred. No. 7.7e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 RPYCWHYPPKPC 20
Db 485 RPYCWHYPPKPC 496

RESULT 3
POLG_HCVJ7 STANDARD; PRT; 737 AA.
AC P27961;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
protein (Envelope protein M); Major envelope protein E; Nonstructural
protein NS1 and NS2] (Fragment).
OS Hepatitis C virus (isolate HC-J7) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11114;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
Tanaka T., Fukuda S., Teuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; D10077; BAA00971.1;
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR PolyPhen; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).

```

```

FT CHAIN 384 733 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 734 >737 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 737 737
SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122F2 CRC64;

Query Match 60.0%; Score 12; DB 1; Length 737;
Best Local Similarity 100.0%; Pred. No. 7.7e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 RPYCWHYPPKPC 20
Db 485 RPYCWHYPPKPC 496

RESULT 4
POLG_HCVBK STANDARD; PRT; 3010 AA.
ID POLG_HCVBK
AC P26663;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome: polyprotein [Contains: Capsid protein C (Core protein) (P22);
Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
(DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
(DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
(DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11105;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91140698; PubMed=1847440;
RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
Onishi E., Andoh T., Yoshida I., Okayama H.;
RT "Structure and organization of the hepatitis C virus genome isolated
RT from human carriers.";
RL J. Virol. 65:1105-1113(1991).
RN [2]
RP SEQUENCE OF 1487-1500.
RX MEDLINE=96235224; PubMed=8647104;
RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetky L.;
RT "Non-structural protein 3 of hepatitis C virus inhibits
RT phosphorylation mediated by cAMP-dependent protein kinase.";
RL Eur. J. Biochem. 237:611-618(1996).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX MEDLINE=97015088; PubMed=8861916;
RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
Moonaw E.W., Adachi T., Hostomsky Z.;
RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT trypsin-like fold and a structural zinc binding site.";
RL Cell 87:331-342(1996).
RN [4]

```


QY 7 DORPYCWHYPP 17
 Db 481 DORPYCWHYPP 491

RESULT 5

POLG_HCVJA STANDARD; PRT; 3010 AA.
 AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Japanese) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91088550; PubMed=2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 RA Sugimura T., Shimotohno K.;
 RT "Molecular cloning of the human hepatitis C virus genome from
 RT Japanese patients with non-A, non-B hepatitis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RN [2]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=91192160; PubMed=1849488;
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K.,
 RA Ohkoshi S., Shimotohno K.;
 RT "Molecular structure of the Japanese hepatitis C viral genome.";
 RL FEBS Lett. 280:325-328(1991).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D90208; BAA14233.1; -
 CC PIR; A39253; GNWVCJ.
 CC HSP; P26663; 1JXP.
 CC MEROPS; S29.001; -
 CC MEROPS; U39.001; -
 CC InterPro; IPR001410; DEAD.
 CC InterPro; IPR002522; HCV capsid.
 CC InterPro; IPR002521; HCV core.
 CC InterPro; IPR002519; HCV env.
 CC InterPro; IPR002531; HCV NS1.
 CC InterPro; IPR002518; HCV NS2.
 CC InterPro; IPR004109; HCV NS3.
 CC InterPro; IPR000745; HCV NS4a.

DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_core; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; Helicase C; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 DR Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 DR Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 KW Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 KW Hepacivirus.
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;
 Query Match 45.0%; Score 9; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.047;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 7 DORPYCWHY 15
 Db 481 DORPYCWHY 489
 RESULT 6
 POLG_HCVJT
 ID POLG_HCVJT STANDARD; PRT; 3010 AA.
 AC Q00269;
 DT 01-APR-1993 (Rel. 25, Created)

01-APR-1993 (Rel. 25, Last sequence update)
 15-SEP-2003 (Rel. 42, Last annotation update)
 Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E2 (GP32) (GP33); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 Hepatitis C virus (isolate HC-JT) (HCV).
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
 NCBI TaxID=31642;
 [1]
 SEQUENCE FROM N.A. PubMed=1318627;
 MEDLINE=92295714; PubMed=1318627;
 Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J., Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 "Molecular cloning of hepatitis C virus genome from a single Japanese carrier: sequence variation within the same individual and among infected individuals."
 Virus Res. 23:39-53(1992).
 -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.
 -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).
 -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

 EMBL; D11168; BAA01943.1; --
 PIR; A45573; A45573.
 PDB; 1AIQ; 25-MAR-98.
 PDB; 1UXP; 14-JAN-98.
 MEROPS; S29.001; --
 MEROPS; U39.001; --
 InterPro; IPR001410; DEAD.
 InterPro; IPR002522; HCV_capsid.
 InterPro; IPR002521; HCV_core.
 InterPro; IPR002519; HCV_env.
 InterPro; IPR002531; HCV_NS1.
 InterPro; IPR002518; HCV_NS2.
 InterPro; IPR004109; HCV_NS3.
 InterPro; IPR000745; HCV_NS4a.
 InterPro; IPR001490; HCV_NS4b.
 InterPro; IPR002868; HCV_NS5a.
 InterPro; IPR002166; HCV_RdRp.
 InterPro; IPR007095; RNA_pol_DS_PS.
 InterPro; IPR007094; RNA_pol_PSVir.
 Pfam; PF01543; HCV_capsid; 1.
 Pfam; PF01542; HCV_core; 1.
 Pfam; PF01539; HCV_env; 1.
 Pfam; PF01560; HCV_NS1; 1.
 Pfam; PF01538; HCV_NS2; 1.
 Pfam; PF02907; HCV_NS3; 1.
 Pfam; PF01006; HCV_NS4a; 1.
 Pfam; PF01001; HCV_NS4b; 1.
 Pfam; PF01506; HCV_NS5a; 1.
 Pfam; PF00271; helicase_C; 1.

DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SMO0487; DEXDc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 NON-STRUCTURAL PROTEIN NS3 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT ACT_SITE 1083 1083 POTENTIAL.
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 1316 1319 ATP (POTENTIAL).
 FT CARBOHYD 196 196 DECH BOX.
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;
 Query Match 45.0%; Score 9; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.047;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 7 DORPCWHY 15
 DB 481 DORPCWHY 489
 RESULT 7
 POLG_HCVTW
 ID POLG_HCVTW STANDARD; PRT; 3010 AA.
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
 OC NCBI TaxID=31645;

RN SEQUENCE FROM N.A.
 RP MEDLINE=92230206; PubMed=1314449;
 RA Chen P.J., Lin M.H., Tai K.P., Liu P.C., Lin C.J., Chen D.S.;
 RT "The Taiwanese hepatitis C virus genome: sequence determination and
 mapping the 5' termini of viral genomic and antigenomic RNA.";
 RL Virology 188:102-113(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M84754; -, NOT_ANNOTATED_CDS.
 DR PIR; A40244; GNMVTV.
 DR PDB; 1N64; 25-FEB-03.
 DR PDB; 1NS3; 08-APR-98.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CORE PROTEIN (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).

FT CHAIN 0 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT ACT_SITE 367 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2329 2329 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;
 Query Match 45.0%; Score 9; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.047;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 DORPYCWHY 15
 DB 484 DORPYCWHY 489
 RESULT 8
 POLG HCVH STANDARD; PRT; 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=1658800;
 RA Inchauspe G., Zebedee S., Lee D.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N.L., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).

[3] X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 MEDLINE=96154321; PubMed=9493270;
 Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 Murcko M.A., Lin C., Caron P.R.;
 "Hepatitis C virus NS3 RNA helicase domain with a bound
 oligonucleotide: the crystal structure provides insights into the mode
 of unwinding";
 Structure 6:89-100(1998).
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position. Cys or Thr in p1 and Ser or Ala in p1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 DR EMBL; M67463; AA45534.1; -.
 DR PIR; A36814; GNWVCH.
 DR PDB; 1HEI; 25-NOV-98.
 DR PDB; 1A1V; 16-FEB-99.
 DR PDB; 1A1R; 17-JUN-98.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR TRANSFAC; T04155; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR007095; RNA pol_DS_PS.
 DR InterPro; IPR007094; RNA pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF002711; helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.

KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 192 383 CAPSID PROTEIN C.
 FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
 FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
 FT CHAIN 810 1026 PROTEIN P7.
 FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
 FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.
 FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4A.
 FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4B.
 FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5A.
 FT CHAIN 3012 369 NONSTRUCTURAL PROTEIN NS5B.
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1224 1226 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1232 1233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1236 1238 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1239 1246 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1247 1248 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1251 1255 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT HELIX 1258 1271 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1272 1272 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1277 1280 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1281 1282 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1283 1285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1291 1295 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT HELIX 1296 1301 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1302 1303 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1312 1316 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1317 1319 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT HELIX 1323 1335 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1336 1340 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1343 1347 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1352 1353 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1360 1361 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1362 1366 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1368 1368 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1373 1375 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1376 1377 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1378 1380 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT HELIX 1382 1385 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1389 1393 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT HELIX 1397 1409 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1410 1411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1414 1417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1419 1420 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1432 1436 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1438 1439 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1450 1453 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1456 1463 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT STRAND 1471 1478
 FT STRAND 1480 1480
 FT HELIX 1481 1488
 FT TURN 1489 1490
 FT STRAND 1497 1501
 FT STRAND 1507 1507
 FT STRAND 1511 1511
 FT HELIX 1514 1527
 FT STRAND 1532 1544
 FT STRAND 1550 1550
 FT HELIX 1555 1564
 FT HELIX 1570 1578
 FT TURN 1579 1580
 FT HELIX 1584 1597
 FT TURN 1598 1598
 FT HELIX 1606 1611
 FT TURN 1614 1618
 FT STRAND 1622 1623
 FT STRAND 1627 1627
 FT STRAND 1635 1636
 FT HELIX 1640 1652
 SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;

Query Match 45.0%; Score 9; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 0.047;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 RPYCWHYPP 17
 |||||
 Db 483 RPYCWHYPP 491

RESULT 9

ID POLG_HCVJ6 STANDARD; PRT; 3033 AA.
 AC P26660;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J6) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92044440; PubMed=158196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 RA Machida A., Miyakawa Y., Mayumi M.;
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 RT from a human carrier; comparison with reported isolates for conserved
 RT and divergent regions.";
 RL J. Gen. Virol. 72:2697-2704 (1991).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Glu in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D00944; BAA00792.1; -;
 DR PIR; J01303; J01303.
 DR HSSP; P27958; 1HEI.
 DR MEROPS; S29.001; -;
 DR INTERPRO; IPR001410; DEAD.
 DR INTERPRO; IPR002522; HCV capsid.
 DR INTERPRO; IPR002521; HCV_core.
 DR INTERPRO; IPR002519; HCV_env.
 DR INTERPRO; IPR002531; HCV_NS1.
 DR INTERPRO; IPR002518; HCV_NS2.
 DR INTERPRO; IPR004109; HCV_NS3.
 DR INTERPRO; IPR000745; HCV_NS4a.
 DR INTERPRO; IPR001490; HCV_NS4b.
 DR INTERPRO; IPR002868; HCV_NS5a.
 DR INTERPRO; IPR002166; HCV RdRP.
 DR INTERPRO; IPR001650; Helicase C.
 DR INTERPRO; IPR007095; RNA pol_Ds_Ps.
 DR INTERPRO; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase C; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 DR Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 DR Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 FT INIT_MBT 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CHAIN 115 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBO


```
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;

Query Match 45.0%; Score 9; DB 1; Length 3033;
Best Local Similarity 100.0%; Pred. No. 0.048;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 RPYCWHYPP 17
Db 485 RPYCWHYPP 493

RESULT 10
POLG HCVJ8
ID POLG HCVJ8 STANDARD; PRT; 3033 AA.
AC P26651.
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J8) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_taxid=11115;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kura K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
RA Fukuda S., Teuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes."
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA) (N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV NS5a.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02307; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral_rdrp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SMO0487; DEXdc; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT TRANSMEM 347 369
FT ACT_SITE 1087 1087
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT NP_BIND 1234 1241
FT SITE 1320 1323
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2359 2359 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;

Query Match 45.0%; Score 9; DB 1; Length 3033;
Best Local Similarity 100.0%; Pred. No. 0.048;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 RPYCWHYPP 17
Db 485 RPYCWHYPP 493
```



```

RESULT 11
POLG_HCVH4      STANDARD;      PRT;      520 AA.
ID  POLG_HCVH4
AC  Q01404;
DT  01-JUL-1993 (Rel. 26, Created)
DT  01-JUL-1993 (Rel. 26, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
DE  Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE  (GP68) (GP70) (NS1) (Fragment).
OS  Hepatitis C virus (isolate HCV-476) (HCV).
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC  Hepacivirus.
OX  NCBI_TaxID=31643;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=93019030; PubMed=1383400;
RA  Abe K., Inchauspe G., Fujisawa K.;
RT  "Genomic characterization and mutation rate of hepatitis C virus
RT  isolated from a patient who contracted hepatitis during an epidemic
RT  of non-A, non-B hepatitis in Japan.";
RL  J. Gen. Virol. 73:2725-2729(1992).
CC  -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC  LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC  PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC  PROTEIN C AND MRNA.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; D10688; BAA01530.1; -
DR  InterPro; IPR002522; HCV_capsid.
DR  InterPro; IPR002521; HCV_core.
DR  InterPro; IPR002531; HCV_Ns1.
DR  Pfam; PF01543; HCV_capsid; 1.
DR  Pfam; PF01542; HCV_core; 1.
DR  Pfam; PF01539; HCV_Ns1; 1.
DR  Pfam; PF01560; HCV_Ns1; 1.
DR  ProDom; PD186062; HCV_Ns1; 1.
KW  Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW  Transmembrane; Nonstructural protein.
FT  INIT_MET 1
FT  CHAIN 1 115
FT  CHAIN 116 191
FT  CHAIN 192 383
FT  CHAIN 384 >520
FT  TRANSMEM 347 369
FT  CARBOHYD 196 196
FT  CARBOHYD 209 209
FT  CARBOHYD 234 234
FT  CARBOHYD 234 234
FT  CARBOHYD 305 305
FT  CARBOHYD 418 418
FT  CARBOHYD 424 424
FT  CARBOHYD 431 431
FT  CARBOHYD 449 449
FT  NON_TER 520 520
SQ  SEQUENCE 520 AA; 56499 MW; AA135246CF20D525 CRC64;

Query Match 40.0%; Score 8; DB 1; Length 520;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 QRPYCWY 15
Db 483 QRPYCWY 490

```

```

RESULT 12
POLG_HCVHK      STANDARD;      PRT;      520 AA.
ID  POLG_HCVHK
AC  Q01403;
DT  01-JUL-1993 (Rel. 26, Created)
DT  01-JUL-1993 (Rel. 26, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
DE  Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE  (GP68) (GP70) (NS1) (Fragment).
OS  Hepatitis C virus (isolate HCV-KF) (HCV).
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC  Hepacivirus.
OX  NCBI_TaxID=31644;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=93019030; PubMed=1383400;
RA  Abe K., Inchauspe G., Fujisawa K.;
RT  "Genomic characterization and mutation rate of hepatitis C virus
RT  isolated from a patient who contracted hepatitis during an epidemic
RT  of non-A, non-B hepatitis in Japan.";
RL  J. Gen. Virol. 73:2725-2729(1992).
CC  -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC  LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC  PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC  PROTEIN C AND MRNA.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; D10687; BAA01529.1; -
DR  PIR; JQ1925; JQ1925.
DR  InterPro; IPR002522; HCV_capsid.
DR  InterPro; IPR002521; HCV_core.
DR  InterPro; IPR002519; HCV_env.
DR  InterPro; IPR002531; HCV_Ns1.
DR  Pfam; PF01543; HCV_capsid; 1.
DR  Pfam; PF01542; HCV_core; 1.
DR  Pfam; PF01539; HCV_Ns1; 1.
DR  Pfam; PF01560; HCV_Ns1; 1.
DR  ProDom; PD186062; HCV_Ns1; 1.
KW  Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW  Transmembrane; Nonstructural protein.
FT  INIT_MET 1
FT  CHAIN 1 115
FT  CHAIN 116 191
FT  CHAIN 192 383
FT  CHAIN 384 >520
FT  TRANSMEM 347 369
FT  CARBOHYD 196 196
FT  CARBOHYD 209 209
FT  CARBOHYD 233 233
FT  CARBOHYD 234 234
FT  CARBOHYD 305 305
FT  CARBOHYD 418 418
FT  CARBOHYD 424 424
FT  CARBOHYD 431 431
FT  CARBOHYD 449 449
FT  NON_TER 520 520
SQ  SEQUENCE 520 AA; 56476 MW; 1D2BD0A6FF27349B CRC64;

Query Match 40.0%; Score 8; DB 1; Length 520;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 QRPYCWY 15

```

```
Db 483 QRPYCHY 490
|||||
RESULT 13
PLR_RANPI STANDARD; PRT; 62 AA.
AC Q90WP7;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Peptide leucine arginine precursor (PLR).
OS Rana pipiens (Northern leopard frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Ranidae; Rana.
OX NCBI_TaxID=8404;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Farragher S., Bjorson A.J., McClean S., Orr D.F., Shaw C.;
RT "Cloning of cDNAs encoding defensive skin secretion peptides from the
RT Northern leopard frog (Rana pipiens).";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Mast cell degranulating peptide.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Skin.
CC -!- SIMILARITY: Belongs to the brevinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ414584; CAC93861.1; -.
CC InterPro; IPR004275; Brevinin.
CC Pfam; PF03032; Brevinin; 1.
CC Amphibian defense peptide; Mast cell degranulation; Amidation; Signal.
FT SIGNAL 1 20
FT PROPEP 21 41
FT PEPTIDE 44 61
FT DISULFID 48 58
FT MOD_RES 61 61
FT SEQUENCE 62 AA; 7113 MW; C8F7F5849A01A1C CRC64;
Query Match 30.0%; Score 6; DB 1; Length 62;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 YPPKPC 20
Db 53 YPPKPC 58
|||||
RESULT 14
EVX2_HETFR STANDARD; PRT; 424 AA.
AC Q91A18;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Homeobox even-skipped homolog protein 2 (EVX-2).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OX NCBI_TaxID=7792;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20144096; PubMed=10677514;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka F., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
RA Kim C.B., Amemiya C., Bailey W., Kawasaki K., Mezey J., Miller W.,
RA Minoshima S., Shimizu N., Wagner G., Ruddle F.;
RT "Hox cluster genomics in the horn shark, Heterodontus francisci.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:1655-1660(2000).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: BELONGS TO THE EVEN-SKIPPED HOMEBOX FAMILY.
CC
CC PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF224263; AAF44638.1; -.
CC HSSP; P14653; 1B72.
CC TRANSFAC; T04488; -.
CC InterPro; IPR001356; Homeobox.
CC Pfam; PF00046; homeobox; 1.
CC ProDom; PD000010; Homeobox; 1.
CC SMART; SM00389; HOX; 1.
CC PROSITE; PS00027; HOMEBOX_1; 1.
CC PROSITE; PS00071; HOMEBOX_2; 1.
CC DNA-BINDING; Developmental protein; Homeobox; Nuclear protein.
FT DNA_BIND 179 238
FT DOMAIN 285 292
FT DOMAIN 340 353
FT SEQUENCE 424 AA; 45629 MW; DA0DAAF3E35D640E CRC64;
SQ
Query Match 30.0%; Score 6; DB 1; Length 424;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GSGPDQ 8
Db 173 GSGPDQ 178
|||||
RESULT 15
POLG_HCVJ2 STANDARD; PRT; 513 AA.
AC P27959;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (Fragment).
OS Hepatitis C virus (isolate HC-J2) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11111;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka F., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

-----
DR EMBL; D10074; BAA00968.1; -
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 >513
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT NON_TER 513 513
SQ SEQUENCE 513 AA; 55704 MW; 943F31E3514CDEF3 CRC64;

```

Query Match 30.0%; Score 6; DB 1; Length 513;
 Best Local Similarity 100.0%; Pred.No. 18;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 PYCWHY 15
 |||||
 Db 484 PYCWHY 489

Search completed: November 21, 2003, 20:59:55
 Job time : 4.85 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:45:57 ; Search time 24.4 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-78

Perfect score: 20

Sequence: 1 ANGSGDQRPYCWHPKPC 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	130	12	Q91ZL6 hepatitis c
2	20	100.0	130	12	Q91ZL3 hepatitis c
3	20	100.0	130	12	Q91ZL5 hepatitis c
4	20	100.0	132	12	Q91Z10 hepatitis c
5	20	100.0	186	12	Q91YP6 hepatitis c
6	20	100.0	186	12	Q91YN2 hepatitis c
7	20	100.0	186	12	Q91YL3 hepatitis c
8	20	100.0	186	12	Q91YQ6 hepatitis c
9	20	100.0	186	12	Q91YP7 hepatitis c
10	20	100.0	186	12	Q91YQ9 hepatitis c
11	20	100.0	186	12	Q91YP8 hepatitis c
12	20	100.0	186	12	Q91YQ5 hepatitis c
13	20	100.0	186	12	Q91YQ7 hepatitis c
14	20	100.0	186	12	Q91YR1 hepatitis c
15	20	100.0	186	12	Q91YP5 hepatitis c
16	20	100.0	186	12	Q91YN3 hepatitis c

17	20	100.0	186	12	Q91YN7 hepatitis c
18	20	100.0	186	12	Q91YQ1 hepatitis c
19	20	100.0	186	12	Q91YP3 hepatitis c
20	20	100.0	186	12	Q91YM9 hepatitis c
21	20	100.0	186	12	Q91YN1 hepatitis c
22	20	100.0	186	12	Q91YQ4 hepatitis c
23	20	100.0	186	12	Q91YQ8 hepatitis c
24	20	100.0	186	12	Q91YN0 hepatitis c
25	20	100.0	186	12	Q91YP2 hepatitis c
26	20	100.0	186	12	Q91YQ2 hepatitis c
27	20	100.0	186	12	Q91YN9 hepatitis c
28	20	100.0	186	12	Q91YP0 hepatitis c
29	20	100.0	186	12	Q91YQ3 hepatitis c
30	20	100.0	186	12	Q91YN7 hepatitis c
31	20	100.0	186	12	Q91YN4 hepatitis c
32	20	100.0	186	12	Q91YP4 hepatitis c
33	20	100.0	186	12	Q91YP1 hepatitis c
34	20	100.0	186	12	Q91YQ8 hepatitis c
35	20	100.0	186	12	Q91YN0 hepatitis c
36	20	100.0	186	12	Q91YN6 hepatitis c
37	20	100.0	186	12	Q91YP9 hepatitis c
38	20	100.0	186	12	Q91YM8 hepatitis c
39	20	100.0	186	12	Q91YN5 hepatitis c
40	20	100.0	186	12	Q91YL4 hepatitis c
41	20	100.0	187	12	Q91YQ6 hepatitis c
42	20	100.0	190	12	Q91YQ2 hepatitis c
43	20	100.0	191	12	Q91YQ6 hepatitis c
44	20	100.0	191	12	Q91YQ3 hepatitis c
45	20	100.0	191	12	Q91YQ8 hepatitis c

ALIGNMENTS

RESULT 1

Q91ZL6 PRELIMINARY; PRT; 130 AA.
ID Q91ZL6;
AC Q91ZL6;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TremBLrel. 22, Last annotation update)
DE Genome polyprotein (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X304C-I;
RX MEDLINE=20173694; PubMed=10708420;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RA Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL; AF230456; AAF68724.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
FT NON_TER 1
FT NON_TER 130 130
SQ SEQUENCE 130 AA; 13899 MW; 0B6A41E9E331F75F CRC64;

Query Match 100.0%; Score 20; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 4.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGDQRPYCWHPKPC 20

DB 103 ANGSGDQRPYCWHPKPC 122

```

RESULT 2
Q91ZL3 PRELIMINARY; PRT; 130 AA.
AC Q91ZL3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]_TaxID=111103;
RP SEQUENCE FROM N.A.
RC STRAIN=X304C-U;
RX MEDLINE=20173694; PubMed=10708420;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RA Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
RT replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL; AF230459; AAF68727.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 130
FT NON_TER 130
SQ SEQUENCE 130 AA; 13827 MW; 0B692289E331F55D CRC64;

Query Match 100.0%; Score 20; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 4.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCYHPPKPC 20
| | | | | | | | | | | | | | | | | |
Db 103 ANGSGPDQRPYCYHPPKPC 122

RESULT 3
Q91ZL5 PRELIMINARY; PRT; 130 AA.
AC Q91ZL5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]_TaxID=111103;
RP SEQUENCE FROM N.A.
RC STRAIN=X304C-K;
RX MEDLINE=20173694; PubMed=10708420;
RA Ray S.C., Mao Q., Lanford R.E., Bassett S., Laeyendecker O.,
RA Wang Y.M., Thomas D.L.;
RT "Hypervariable region 1 sequence stability during hepatitis C virus
RT replication in chimpanzees.";
RL J. Virol. 74:3058-3066(2000).
DR EMBL; AF230457; AAF68725.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 130
FT NON_TER 130
SQ SEQUENCE 130 AA; 13827 MW; 0B692289E331F55D CRC64;

Query Match 100.0%; Score 20; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 4.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCYHPPKPC 20
| | | | | | | | | | | | | | | | | |
Db 103 ANGSGPDQRPYCYHPPKPC 122

RESULT 4
Q81210 PRELIMINARY; PRT; 132 AA.
AC Q81210;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Envelope polyprotein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]_TaxID=111103;
RP SEQUENCE FROM N.A.
RC STRAIN=SW14;
RA Johanson B., Yun Z.-B., Staland A., Sonnerborg A.;
RT "Genotyping of Swedish hepatitis C virus using E2/NS1 sequences.";
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L16890; AAA45530.2; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT CHAIN 1 >12 E1 GLYCOPROTEIN.
FT CHAIN 13 >132 E2/NS1 GLYCOPROTEIN.
FT NON_TER 132
FT NON_TER 132
SQ SEQUENCE 132 AA; 14340 MW; BD0808EA0664CACC CRC64;

Query Match 100.0%; Score 20; DB 12; Length 132;
Best Local Similarity 100.0%; Pred. No. 4.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCYHPPKPC 20
| | | | | | | | | | | | | | | | | |
Db 104 ANGSGPDQRPYCYHPPKPC 123

RESULT 5
Q91YP6 PRELIMINARY; PRT; 186 AA.
AC Q91YP6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]_TaxID=111103;
RP SEQUENCE FROM N.A.
RC STRAIN=P4;
RX MEDLINE=20230065; PubMed=10764648;
RA Farci P., Shimoda A., Coliana A., Diaz G., Peddis G., Melpolder J.C.,
RA Strazzer A., Chien D.Y., Munoz S.J., Balestrieri A., Purcell R.H.,
RA Alter H.J.;
RT "The outcome of acute hepatitis C predicted by the evolution of the
RT viral quasispecies.";
RL Science 288:339-344(2000).
DR EMBL; AF245802; AAF65119.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.

```

```
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 19846 MW; 3E51A0D1F8B9D09B CRC64;

Query Match 100.0%; Score 20; DB 12; Length 186;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20
Db 158 ANGSGPDQRPYCWHPKPC 177

RESULT 6
ID Q9IYN2 PRELIMINARY; PRT; 186 AA.
AC Q9IYN2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P4;
RX MEDLINE=20230065; PubMed=10764648;
RA Farci P., Shimoda A., Coiana A., Diaz G., Peddis G., Melpolder J.C.,
RA Strazzera A., Chien D.Y., Munoz S.J., Balestrieri A., Purcell R.H.,
RA Alter H.J.;
RT "The outcome of acute hepatitis C predicted by the evolution of the
RT viral quasispecies.";
RL Science 288:339-344(2000).
DR EMBL; AF245816; AAF65133.1; -.
DR InterPro; IPR002519; HCV env.
DR Pfam; PF01539; HCV env; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 19813 MW; 62A5327E5439496F CRC64;

Query Match 100.0%; Score 20; DB 12; Length 186;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20
Db 158 ANGSGPDQRPYCWHPKPC 177

RESULT 7
ID Q9IYL3 PRELIMINARY; PRT; 186 AA.
AC Q9IYL3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P4;
RX MEDLINE=20230065; PubMed=10764648;
RA Farci P., Shimoda A., Coiana A., Diaz G., Peddis G., Melpolder J.C.,
RA Strazzera A., Chien D.Y., Munoz S.J., Balestrieri A., Purcell R.H.,
RA Alter H.J.;
RT "The outcome of acute hepatitis C predicted by the evolution of the
RT viral quasispecies.";
RL Science 288:339-344(2000).
DR EMBL; AF245816; AAF65133.1; -.
DR InterPro; IPR002519; HCV env.
DR Pfam; PF01539; HCV env; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 19813 MW; 62A5327E5439496F CRC64;

Query Match 100.0%; Score 20; DB 12; Length 186;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20
Db 158 ANGSGPDQRPYCWHPKPC 177

RESULT 8
ID Q9IYQ6 PRELIMINARY; PRT; 186 AA.
AC Q9IYQ6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P4;
RX MEDLINE=20230065; PubMed=10764648;
RA Farci P., Shimoda A., Coiana A., Diaz G., Peddis G., Melpolder J.C.,
RA Strazzera A., Chien D.Y., Munoz S.J., Balestrieri A., Purcell R.H.,
RA Alter H.J.;
RT "The outcome of acute hepatitis C predicted by the evolution of the
RT viral quasispecies.";
RL Science 288:339-344(2000).
DR EMBL; AF245792; AAF65109.1; -.
DR InterPro; IPR002519; HCV env.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 19699 MW; 464557835DA4C5F9 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 186;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20
Db 158 ANGSGPDQRPYCWHPKPC 177
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=P4;
RX MEDLINE=20230065; PubMed=10764648;
RA Farci P., Shimoda A., Coiana A., Diaz G., Peddis G., Melpolder J.C.,
RA Strazzera A., Chien D.Y., Munoz S.J., Balestrieri A., Purcell R.H.,
RA Alter H.J.;
RT "The outcome of acute hepatitis C predicted by the evolution of the
RT viral quasispecies.";
RL Science 288:339-344(2000).
DR EMBL; AF245835; AAF65152.1; -.
DR InterPro; IPR002519; HCV env.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 19699 MW; 5D4557835DA4C5FD CRC64;

Query Match 100.0%; Score 20; DB 12; Length 186;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20
Db 158 ANGSGPDQRPYCWHPKPC 177

RESULT 8
ID Q9IYQ6 PRELIMINARY; PRT; 186 AA.
AC Q9IYQ6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P4;
RX MEDLINE=20230065; PubMed=10764648;
RA Farci P., Shimoda A., Coiana A., Diaz G., Peddis G., Melpolder J.C.,
RA Strazzera A., Chien D.Y., Munoz S.J., Balestrieri A., Purcell R.H.,
RA Alter H.J.;
RT "The outcome of acute hepatitis C predicted by the evolution of the
RT viral quasispecies.";
RL Science 288:339-344(2000).
DR EMBL; AF245792; AAF65109.1; -.
DR InterPro; IPR002519; HCV env.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 19699 MW; 464557835DA4C5F9 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 186;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20
Db 158 ANGSGPDQRPYCWHPKPC 177
```

```
RESULT 9
Q9IYP7 PRELIMINARY; PRT; 186 AA.
ID Q9IYP7;
AC Q9IYP7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P4;
RX MEDLINE=20230065; PubMed=10764648;
RA Farci P., Shimoda A., Coiana A., Diaz G., Peddis G., Melpolder J.C.,
RA Strazzera A., Chien D.Y., Munoz S.J., Balestrieri A., Purcell R.H.,
RA Alter H.J.;
RT "The outcome of acute hepatitis C predicted by the evolution of the
RT viral quasispecies.";
RL Science 288:339-344(2000).
DR EMBL; AF245801; AAF65118.1; -.
DR InterPro; IPR002519; HCV env.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 19799 MW; 0041CF2070AE10F0 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 186;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20
DB 158 ANGSGPDQRPYCWHPKPC 177

RESULT 10
Q9IYQ9 PRELIMINARY; PRT; 186 AA.
ID Q9IYQ9;
AC Q9IYQ9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P4;
RX MEDLINE=20230065; PubMed=10764648;
RA Farci P., Shimoda A., Coiana A., Diaz G., Peddis G., Melpolder J.C.,
RA Strazzera A., Chien D.Y., Munoz S.J., Balestrieri A., Purcell R.H.,
RA Alter H.J.;
RT "The outcome of acute hepatitis C predicted by the evolution of the
RT viral quasispecies.";
RL Science 288:339-344(2000).
DR EMBL; AF245789; AAF65106.1; -.
DR InterPro; IPR002519; HCV env.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 19799 MW; 0041CF2070AE10F0 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 186;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20
DB 158 ANGSGPDQRPYCWHPKPC 177
```

```
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 19809 MW; B65AA07568CC6C99 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 186;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20
DB 158 ANGSGPDQRPYCWHPKPC 177

RESULT 11
Q9IYP8 PRELIMINARY; PRT; 186 AA.
ID Q9IYP8;
AC Q9IYP8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P4;
RX MEDLINE=20230065; PubMed=10764648;
RA Farci P., Shimoda A., Coiana A., Diaz G., Peddis G., Melpolder J.C.,
RA Strazzera A., Chien D.Y., Munoz S.J., Balestrieri A., Purcell R.H.,
RA Alter H.J.;
RT "The outcome of acute hepatitis C predicted by the evolution of the
RT viral quasispecies.";
RL Science 288:339-344(2000).
DR EMBL; AF245800; AAF65117.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 19737 MW; 025C1718A90DDA33 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 186;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCWHPKPC 20
DB 158 ANGSGPDQRPYCWHPKPC 177

RESULT 12
Q9IYQ5 PRELIMINARY; PRT; 186 AA.
ID Q9IYQ5;
AC Q9IYQ5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P4;
RX MEDLINE=20230065; PubMed=10764648;
RA Farci P., Shimoda A., Coiana A., Diaz G., Peddis G., Melpolder J.C.,
```



```

RA Strazzera A., Chien D.Y., Munoz S.J., Balestrieri A., Purcell R.H.,
RA Alter H.J.;
RT "The outcome of acute hepatitis C predicted by the evolution of the
RT viral quasispecies.";
RL Science 288:339-344(2000).
DR EMBL; AF245793; AAF65110.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 19900 MW; 7991A0D1F8A24662 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 186;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCHWYPPKPC 20
Db |||||
158 ANGSGPDQRPYCHWYPPKPC 177

RESULT 13
Q9IYQ7 PRELIMINARY; PRT; 186 AA.
AC Q9IYQ7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polypeptide (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RC STRAIN=P4;
RX MEDLINE=2030065; PubMed=10764648;
RA Farci P., Shimoda A., Coiana A., Diaz G., Peddis G., Melpolder J.C.,
RA Strazzera A., Chien D.Y., Munoz S.J., Balestrieri A., Purcell R.H.,
RA Alter H.J.;
RT "The outcome of acute hepatitis C predicted by the evolution of the
RT viral quasispecies.";
RL Science 288:339-344(2000).
DR EMBL; AF245791; AAF65108.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 19906 MW; 3E51A0D1E5852349 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 186;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCHWYPPKPC 20
Db |||||
158 ANGSGPDQRPYCHWYPPKPC 177

RESULT 14
Q9IYR1 PRELIMINARY; PRT; 186 AA.
AC Q9IYR1;

```

```

DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polypeptide (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RC STRAIN=P4;
RX MEDLINE=2030065; PubMed=10764648;
RA Farci P., Shimoda A., Coiana A., Diaz G., Peddis G., Melpolder J.C.,
RA Strazzera A., Chien D.Y., Munoz S.J., Balestrieri A., Purcell R.H.,
RA Alter H.J.;
RT "The outcome of acute hepatitis C predicted by the evolution of the
RT viral quasispecies.";
RL Science 288:339-344(2000).
DR EMBL; AF245787; AAF65104.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 19846 MW; 3E51A0D1F8B9D09B CRC64;

Query Match 100.0%; Score 20; DB 12; Length 186;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGSGPDQRPYCHWYPPKPC 20
Db |||||
158 ANGSGPDQRPYCHWYPPKPC 177

RESULT 15
Q9IYPS PRELIMINARY; PRT; 186 AA.
AC Q9IYPS;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polypeptide (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RC STRAIN=P4;
RX MEDLINE=2030065; PubMed=10764648;
RA Farci P., Shimoda A., Coiana A., Diaz G., Peddis G., Melpolder J.C.,
RA Strazzera A., Chien D.Y., Munoz S.J., Balestrieri A., Purcell R.H.,
RA Alter H.J.;
RT "The outcome of acute hepatitis C predicted by the evolution of the
RT viral quasispecies.";
RL Science 288:339-344(2000).
DR EMBL; AF245803; AAF65120.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 19846 MW; 3E51A0D1F8B9D09B CRC64;

```

Query Match 100.0%; Score 20; DB 12; Length 186;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANSGGDQRPYCWHPKPC 20
| | | | | | | | | | | | | | | | | | | | | |
Db 158 ANSGGDQRPYCWHPKPC 177

Search completed: November 21, 2003, 21:08:16
Job time : 25.45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:22 ; Search time 32.1 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-79

Perfect score: 20

Sequence: 1 WHYPKPGCIVPAKSVCGPV 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	16	AA1984508
2	20	100.0	20	17	AA1991002
3	20	100.0	20	23	AA198708
4	20	100.0	30	23	AA1984630
5	20	100.0	42	13	AA198751
6	20	100.0	44	16	AA1984237
7	20	100.0	76	21	AA198528
8	20	100.0	144	14	AA1983998
9	20	100.0	174	14	AA1984441

10	20	100.0	192	19	AA1987009
11	20	100.0	254	22	AA1988043
12	20	100.0	278	14	AA1983997
13	20	100.0	319	15	AA1985330
14	20	100.0	363	22	AA1988042
15	20	100.0	402	14	AA1984438
16	20	100.0	454	10	AA1980183
17	20	100.0	454	10	AA1982049
18	20	100.0	454	21	AA198526
19	20	100.0	463	14	AA1983588
20	20	100.0	463	14	AA1983187
21	20	100.0	463	22	AA1989007
22	20	100.0	480	14	AA1983992
23	20	100.0	513	13	AA1984086
24	20	100.0	527	13	AA1985136
25	20	100.0	531	22	AA1982622
26	20	100.0	621	14	AA1983185
27	20	100.0	622	14	AA1983591
28	20	100.0	622	22	AA1989010
29	20	100.0	637	24	AA1987410
30	20	100.0	663	17	AA1982935
31	20	100.0	663	20	AA1987615
32	20	100.0	733	14	AA1988278
33	20	100.0	738	14	AA1983592
34	20	100.0	738	22	AA1989011
35	20	100.0	2010	23	AA1984801
36	20	100.0	2435	13	AA1985135
37	20	100.0	2436	10	AA1982050
38	20	100.0	2436	10	AA1980288
39	20	100.0	2436	13	AA1988582
40	20	100.0	2772	11	AA1988123
41	20	100.0	2772	21	AA198540
42	20	100.0	2816	14	AA1984009
43	20	100.0	2894	13	AA1984440
44	20	100.0	2894	15	AA1980230
45	20	100.0	2955	11	AA1988124

ALIGNMENTS

RESULT 1

AA1984508
ID AA1984508 standard; peptide; 20 AA.
AC AA1984508;
XX
DT 06-JAN-1997 (first entry)
XX
DE Hepatitis C virus peptide NS1-7 (residues 487-506).
XX
KW Hepatitis C virus; HCV; immunogen; non-structural region; NS1;
KW immunodominant; T cell epitope; vaccine.
XX
OS Hepatitis C virus.

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

HCV nucleocapsid c
Amino acid sequenc
Th E2/NS1 protein.
Anti-HCV antibody
Amino acid sequenc
Sequence of glycop
Sequence of hepati
Sequence encoded b
Protein encoded by
HCV CKS-NS1S1 fusi
Sequence of subfra
HCV recombinant an
HCV-1 E2/NS1 prote
NANB hepatitis vir
HCV polypeptide 2.
Chimeric HCV E2661
Sequence of subfra
HCV CKS-NS1S1-NS1S
HCV recombinant an
Hepatitis C virus
HCV1 E2 + NS2 poly
Hepatitis C virus
NANB hepatitis vir
HCV CKS-full lengt
HCV recombinant an
HCV HepC cassette
HCV polypeptide 1.
Sequence encoded i
Peptide encoded by
HCV amino acid seq
Hepatitis C virus
Protein encoded by
HCV-1 polypeptide.
Composite HCV HC-J
Composite hepatiti
Hepatitis C virus

OS
PN WO9512677-A2.
XX
PD 11-MAY-1995.
XX
PF 28-OCT-1994; 94WO-EP03555.
XX
PR 04-NOV-1993; 93EP-0402718.
XX
(INNO-) INNOGENETICS NV.
PI Deleys R, Leroux-Roels G, Maertens G;
DR WPI; 1995-193822/25.
XX
PT Hepatitis C Virus immunogenic polypeptide contg. a T-cell
PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in
PT production of vaccines, therapeutic agents, etc.

```
XX Example 4; Page 51; 105pp; English.
PS
XX A series of overlapping peptides (including the present sequence) was
CC synthesised based on sequences in the core, E1 and E2/NS1 regions of
CC hepatitis C virus. The peptides were used as antigens in lympho-
CC proliferative assays to identify the main T-cell epitopes.
XX
SQ Sequence 20 AA;
Query Match 100.0%; Score 20; DB 16; Length 20;
Best Local Similarity 100.0%; Pred. No. 3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WHYPKPCGIVPAKSVCGPV 20
DB 1 WHYPKPCGIVPAKSVCGPV 20
|||||
RESULT 2
AAR91002
ID AAR91002 standard; peptide; 20 AA.
XX
AC AAR91002;
XX
DT 25-SEP-1996 (first entry)
XX
DE HCV E2 peptide E2-7B for competition studies.
XX
KW HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;
KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.
XX
OS Synthetic.
XX
PN WO9604385-A2.
XX
PD 15-FEB-1996.
XX
PF 31-JUL-1995; 95WO-EP03031.
XX
PR 29-JUL-1994; 94EP-0870132.
XX
PA (INNO-) INNOGENETICS NV.
XX
PI Bosman F, Buyse M, De Martynoff G, Maertens G;
XX
DR WPI; 1996-129401/13.
XX
PT Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope
PT proteins - in presence of disulphide bond cleavage agent, to
PT produce proteins suitable for direct use in vaccines or diagnostic
PT assays of HCV
XX
XX Example 7; Page 67; 146pp; English.
XX
CC AAR90965-R90980 and AAR90995-R91015 represent synthetic hepatitis C
CC virus (HCV) E1 and E2 peptides used in competition studies. This
CC sequence represents a synthetic E2 peptide, and corresponds to residues
CC 487-506 of the E2 protein sequence. These sequences are useful for in
CC vitro monitoring of HCV disease, or prognosis of the response to
CC interferon treatment of patients suffering from HCV infection. These
CC sequences compete with the proteins produced by AAT12704-T12709 and
CC AAT12961-T12974, which are included in vectors for the production of
CC recombinant E1, E2, and E1/E2. The recombinant proteins can then be
CC isolated and purified by carrying out a disulphide bond cleavage, or a
CC reduction step with a disulphide bond cleavage agent, after lysis of
CC recombinant host cells. The constructs containing the purified HCV
CC envelope proteins can be used for vaccinating humans against HCV, for in
CC vitro detection of HCV antibodies in a sample, and in a serotyping assay
CC for detecting one or more serological types of HCV present in a
CC biological sample. The constructs can also be immobilised on a solid
CC substrate and incorporated into a reversed phase hybridisation assay for
CC determining the presence or the genotype of HCV. The new purification
method preserves the conformation of the recombinantly expressed E1, E2
and E1/E2, and eliminates contaminating proteins. Antigens isolated
using this method are more reactive with human sera than those isolated
by known techniques.
XX
SQ Sequence 20 AA;
Query Match 100.0%; Score 20; DB 17; Length 20;
Best Local Similarity 100.0%; Pred. No. 3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WHYPKPCGIVPAKSVCGPV 20
DB 1 WHYPKPCGIVPAKSVCGPV 20
|||||
RESULT 3
AAO18708
ID AAO18708 standard; Peptide; 20 AA.
XX
AC AAO18708;
XX
DT 24-OCT-2002 (first entry)
XX
DE Hepatitis C virus E2 protein derived peptide E2-7B.
XX
KW Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide;
KW immunostimulant; vaccine.
XX
OS Hepatitis C virus.
XX
PN WO200255548-A2.
XX
PD 18-JUL-2002.
XX
PF 11-JAN-2002; 2002WO-EP00219.
XX
PR 11-JAN-2001; 2001US-260699P.
XX
PR 30-AUG-2001; 2001US-315768P.
XX
PA (INNO-) INNOGENETICS NV.
XX
PI Maertens G, Bosman F, Buyse M;
XX
DR WPI; 2002-599657/64.
XX
PT New therapeutic vaccine compositions comprising at least one purified
PT recombinant hepatitis C virus (HCV) single or specific oligomeric
PT recombinant envelope protein E1 or E2, useful for immunizing humans
XX from HCV infection
XX
XX Example 7; Page 227; 243pp; English.
XX
CC The present invention relates to new therapeutic vaccine compositions for
CC inducing hepatitis C virus (HCV)-specific antibodies, comprising a
CC composition containing at least one purified recombinant HCV single or
CC specific oligomeric recombinant envelope proteins selected from an E1 and
CC an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are
CC useful for inducing HCV-specific antibodies or for immunising humans
CC against HCV. The recombinant HCV E1 and/or E2 proteins are useful as
CC vaccines or therapeutics, in HCV screening and confirmatory antibody
CC tests, for raising antibodies, in the preparation of medicament, and for
CC in vitro monitoring of HCV disease or prognosing the response to
CC treatment of patients suffering from HCV infection. The present sequence
CC is a peptide derived from the proteins of the invention.
XX
SQ Sequence 20 AA;
Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WHYPKPCGIVPAKSVCGPV 20
```

Db 1 WHYPKPCGIVPAKSVCGPV 20
|||||

RESULT 4
AAU84630
ID AAU84630 standard; Peptide; 30 AA.
XX AC AAU84630;

DT 08-MAY-2002 (first entry)

DE HCV HepC1a segment 33.

XX Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
KW viral infection; human immunodeficiency virus; melanoma;
KW bacterial infection; Salmonella; Legionella; parasitic infection;
KW Trypanosoma; Toxoplasma; Giardia.

XX Hepatitis C virus.

XX WO200190197-A1.

XX 29-NOV-2001.

XX 25-MAY-2001; 2001WO-AU00622.

XX 26-MAY-2000; 2000AU-0007761.

XX (AUSU) UNIV AUSTRALIAN NAT.

XX Thomson SA, Ramshaw IA;

XX WPI: 2002-147575/19.

XX N-PSDB; ABK36468.

XX New synthetic polypeptides having several different segments of at
PT least one parent polypeptide linked together differently compared to
PT the linkage in the parent polypeptide, for inducing immune response
PT against a pathogen or cancer.

XX Example 2; Fig 26; 364pp; English.

XX The invention relates to a new synthetic polypeptide (I) comprising
CC several different segments of at least one parent polypeptide linked
CC together in a different relationship relative to their linkage in the
CC parent polypeptide to impede, abrogate or otherwise alter at least one
CC function associated with the parent polypeptide and for inducing an
CC immune response against a pathogen or cancer. Also included are a
CC synthetic polynucleotide encoding and a computer system for
CC designing the synthetic polypeptides. The synthetic polypeptides and
CC polynucleotides are referred to as a vaccine. The synthetic polypeptide is
CC useful for modulating immune responses preferably directed against a
CC pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix,
CC colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone
CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.
CC Compositions comprising the polypeptide may be used in the treatment or
CC prophylaxis against viral (such as infections caused by HIV (human
CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
CC (e.g., infections caused by Neisseria, Meningococcus, Haemophilus,
CC Salmonella, Streptococcus, Legionella and Mycobacterium or parasitic
CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
CC Trypanosoma, Toxoplasma and Giardia) infections. The present
CC sequence is a peptide derived from a parent protein used to
CC construct a vaccine of the invention.

XX Sequence 30 AA;

Query Match 100.0%; Score 20; DB 23; Length 30;
Best Local Similarity 100.0%; Pred. No. 4.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WHYPKPCGIVPAKSVCGPV 20
|||||
Db 9 WHYPKPCGIVPAKSVCGPV 28

RESULT 5
AAR20751
ID AAR20751 standard; peptide; 42 AA.
XX AC AAR20751;

XX 25-MAR-2003 (updated)

DT 05-MAY-1992 (first entry)

XX Peptide 1 based on immunoreactive region of Hepatitis C virus.

XX Non-A, non-B hepatitis virus; non-structural protein; vaccine.

XX Synthesis.

XX EP468527-A.

XX 29-JAN-1992.

XX 26-JUL-1991; 91EP-0112620.

XX 24-JUN-1991; 91US-0719819.

XX 26-JUL-1990; 90US-0558799.

XX 07-FEB-1991; 91US-0651735.

XX 11-MAR-1991; 91US-0667275.

XX (UNBI-) UNITED BIOMEDICAL INC.

XX Chang YW, Hosein B;

XX WPI: 1992-034279/05.

XX New synthetic peptide specific for HCV antibodies - for detection
PT of HCV or NANBHV e.g. by enzyme-linked immunosorbent assay and is
PT immunogen for preparation of vaccines

XX Claim 1; Page 88; 98pp; English.

XX This peptide is one of 19 specifically claimed antigens based on the
CC immunoreactive regions of the envelope protein and non-structural
CC proteins NS-1, NS-2, NS-3 and NS-5 for the Hepatitis C virus. The
CC invention also concerns analogues, segments, mixtures, conjugates
CC and polymers of these peptides. The C-terminal amino acid may be
CC amidated. See also AAR20752-R20782.
CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 42 AA;

Query Match 100.0%; Score 20; DB 13; Length 42;
Best Local Similarity 100.0%; Pred. No. 5.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 WHYPKPCGIVPAKSVCGPV 20
|||||
Db 21 WHYPKPCGIVPAKSVCGPV 40

RESULT 6
AAR74237
ID AAR74237 standard; Peptide; 44 AA.
XX AC AAR74237;

XX 25-MAR-2003 (updated)

DT 24-DEC-1995 (first entry)

XX HCV antigenic NS1 peptide SSAL.

KW Structured antigenic peptide library; SSAL; vaccine; diagnostic;
 KW therapeutic; hepatitis C virus, HCV.

OS Synthetic.

XX Key Location/Qualifiers

PH Misc-difference 1 /note= "Q8;I1;V1"

FT Misc-difference 5 /note= "P8;T1;A1"

FT Misc-difference 6 /note= "I8;I2"

FT Misc-difference 7 /note= "S5;T3;E1;Q1"

FT Misc-difference 8 /note= "Y9;H1"

FT Misc-difference 9 /note= "A6;D1;T1;E2"

FT Misc-difference 10 /note= "N5;E2;M1;T1;D1"

FT Misc-difference 11 /note= "unknown8;N2"

FT Misc-difference 12 /note= "unknown8;V2"

FT Misc-difference 13 /note= "G5;S1;P1;A1;T2"

FT Misc-difference 14 /note= "S6;E1;D1;N2"

FT Misc-difference 15 /note= "G5;R1;S1;I1;D1;P1"

FT Misc-difference 16 /note= "P4;I1;S2;Q1;G1;E1"

FT Misc-difference 18 /note= "Q7;E1;N2"

FT Misc-difference 26 /note= "P8;A2"

FT Misc-difference 28 /note= "K4;R5;P1"

FT Misc-difference 29 /note= "P8;Q2"

FT Misc-difference 31 /note= "G9;T1"

FT Misc-difference 32 /note= "I9;V1"

FT Misc-difference 34 /note= "P9;S1"

FT Misc-difference 36 /note= "K5;S4;R1"

FT Misc-difference 37 /note= "S6;E1;Q2;T1"

FT WO9511998-A1.

XX 04-MAY-1995.

XX 26-OCT-1994; 94WO-US12268.

XX 26-OCT-1993; 93US-0143412.

XX (UNBI-) UNITED BIOMEDICAL INC.

XX Wang CY, Zamb TJ, Ye J, Kaminsky SM, Hosein B, Nixon DF, Koff CW;

PI Kowalski J, Walfield AM;

XX WPI; 1995-178890/23.

XX Structured antigenic peptide libraries contain some invariant amino

PT acids - accommodate variations in antigenic structure so are

PT effective against many different strains of e.g. rapidly mutating

XX viruses in vaccines

XX Claim 4; Page 91; 216pp; English.

CC In a structured synthetic antigen library, specific amino acids and
 CC their frequency of appearance at a variant locus within aligned
 CC peptide sequences are defined by the primary sequences of the several
 CC variants that make up the alignment used to construct the antigen
 CC peptide library. Branched SSALs representing the highly variable
 CC yet antigenic regions of the HCV Env and NS1 proteins are given in
 CC AAR74233-37 and were used as the key ingredients in a polyvalent
 CC vaccine for highly divergent global strains of HCV.
 CC (Updated on 25-MAR-2003 to correct PI field.)

XX SQ Sequence 44 AA;

Query Match 100.0%; Score 20; DB 16; Length 44;

Best Local Similarity 100.0%; Pred. No. 5.3e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGPV 20

Db 23 WHYPKPCGIVPAKSVCGPV 42

RESULT 7

AAB18528

ID AAB18528 standard; Protein; 76 AA.

XX AC AAB18528;

XX DT 15-JAN-2001 (first entry)

XX DE Protein encoded by a novel hepatitis C virus cDNA clone 26j.

XX KW Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;

XX KW viral infectivity; viral replication.

XX OS Hepatitis C virus.

XX PN EP1034785-A2.

XX PD 13-SEP-2000.

XX PF 16-MAR-1990; 2000EP-0109602.

XX PR 17-MAR-1989; 89US-0325338.

XX PR 20-APR-1989; 89US-0341334.

XX PR 18-MAY-1989; 89US-0355002.

XX PR 16-MAR-1990; 90EP-0302866.

XX PA (CHIR) CHIRON CORP.

XX PI Houghton M, Choo Q, Kuo G;

XX DR WPI; 2000-566891/53.

XX DR N-PSDB; AAA75284.

XX Novel composition comprising a hepatitis C virus antisense

PT polynucleotide which is complementary to or corresponds to a sense

PT strand of the virus genome, and selectively hybridises to it -

XX Example; Fig 4; 75pp; English.

XX The specification describes a pharmaceutical composition which

CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The

CC HCV is characterized by a positive stranded RNA genome which has

CC 40% homology at the polypeptide level to a HCV polyprotein. The

CC antisense polynucleotide binds to cellular polynucleotides which

CC enhance and/or are required for viral infectivity, replicative

CC ability or chronicity. The antisense polynucleotides may also be

CC designed to bind with high specificity, to be of increased stability,

CC to be stable and to have low toxicity. The composition also comprises

CC an agent which causes viral RNA to be inactive. The composition

CC is used for preventing HCV replication in a system. The present

CC sequence is encoded by a novel HCV cDNA sequence, which is used in the

CC course of the invention.

```

XX SQ Sequence 76 AA;
Query Match 100.0%; Score 20; DB 21; Length 76;
Best Local Similarity 100.0%; Pred. No. 7.9e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPGIVPAKSVCGPV 20
Db 47 WHYPPKPGIVPAKSVCGPV 66
|||||

RESULT 8
AAR33998
ID AAR33998 standard; Protein; 144 AA.
XX AC AAR33998;
XX DT 25-MAR-2003 (updated)
DT 26-JUL-1993 (first entry)
XX DE HC-J1 E2/NS1 protein.
XX KW Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
KW domain; immunological; cross-reactive; envelope protein; vaccine;
KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.
XX OS Synthetic.
XX PN WO9306126-A1.
XX PD 01-APR-1993.
XX PF 11-SEP-1992; 92WO-US07683.
XX PR 13-SEP-1991; 91US-0759575.
XX PA (CHIR ) CHIRON CORP.
XX PI Houghton M, Weiner AJ;
XX WPI; 1993-117468/14.
XX DR Immuno-reactive hepatitis C virus polypeptide compans. - contg.
PT at least 2 sequences from the first variable domain of distinct
PT HCV isolates
XX PS Disclosure; Fig 3; 106pp; English.
XX The sequences given in AAR33998-002 represent a portion of the E2/NS1
CC protein encoded by group I and group II HCV isolates, from amino acid
CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
CC 30 amino acids which shows large variation between nearly all isolates.
CC This is an important immunoreactive domain. This putative envelope
CC glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera
CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
CC flaviviruses, both of which confer protective immunity in hosts
CC vaccinated with these polypeptides. It has been discovered that a
CC number of important HCV epitopes vary among viral isolates and that
CC these epitopes can be mapped to specific domains. This meant that
CC immunologically cross-reactive polypeptides which focus on variable
CC rather than constant domains can be produced. See also AAR339134-48
CC and AAR33982-91.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 144 AA;
Query Match 100.0%; Score 20; DB 14; Length 144;
Best Local Similarity 100.0%; Pred. No. 1.2e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPGIVPAKSVCGPV 20
Db 148 WHYPPKPGIVPAKSVCGPV 167
|||||

RESULT 10
AAW67009
ID AAW67009 standard; protein; 192 AA.
XX AC AAW67009;
XX DT 02-MAR-1999 (first entry)
XX DE HCV nucleocapsid core protein.

```

```

QY 1 WHYPPKPGIVPAKSVCGPV 20
Db 118 WHYPPKPGIVPAKSVCGPV 137
|||||

RESULT 9
AAR34441
ID AAR34441 standard; Protein; 174 AA.
XX AC AAR34441;
XX DT 25-MAR-2003 (updated)
DT 09-AUG-1993 (first entry)
XX DE Sequence of glycoprotein E2/NS1 in clone J1(JM).
XX KW Hepatitis C virus; envelope protein; glycoprotein; E2/NS1;
KW diagnostic reagent.
XX OS Hepatitis C virus.
XX PN EP537626-A1.
XX PD 21-APR-1993.
XX PF 08-OCT-1992; 92EP-0117191.
XX PR 08-OCT-1991; 91JP-0260824.
XX PA (NAHE-)o NAT INST OF HEALTH.
XX PI Harada S, Honda Y, Miyamura T, Saito I;
XX WPI; 1993-127516/16.
XX DR N-PSDB; AAQ40333.
XX PT Diagnostic reagent for hepatitis C virus - comprises second
PT envelope protein or first non-structural protein encoded by HCV
PT gene and has sugar chain
XX PS Claim 2; Pages 40-41; 58pp; English.
XX CC Glycoprotein E2/NS1 is derived from the second envelope protein or
CC first non-structural protein encoded by the genome of HCV. The
CC nucleic acid is extracted from the serum of the patient of hepatitis
CC C. The serum is pref. mixed with transfer RNA (tRNA) as a carrier
CC of virus RNA. As a technique of cloning cDNA from the nucleic acid,
CC it is preferred to use polymerase chain reaction method. In the
CC reaction, any commercially available random primers or synthesized
CC DNA having a base sequence similar to that of primer AS1 may be used
CC as a primer. Representative examples of sense primers include S1.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 174 AA;
Query Match 100.0%; Score 20; DB 14; Length 174;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPGIVPAKSVCGPV 20
Db 148 WHYPPKPGIVPAKSVCGPV 167
|||||

RESULT 10
AAW67009
ID AAW67009 standard; protein; 192 AA.
XX AC AAW67009;
XX DT 02-MAR-1999 (first entry)
XX DE HCV nucleocapsid core protein.

```


XX The sequences given in AAR33992-002 represent a portion of the E2/NS1
CC protein encoded by group I and group II HCV isolates, from amino acid
CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
CC 30 amino acids which shows large variation between nearly all isolates.
CC This is an important immunoreactive domain. This putative envelope
CC glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera
CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
CC flaviviruses, both of which confer protective immunity in hosts
CC vaccinated with these polypeptides. It has been discovered that a
CC number of important HCV epitopes vary among viral isolates and that
CC these epitopes can be mapped to specific domains. This meant that
CC immunologically cross-reactive polypeptides which focus on variable
CC rather than constant domains can be produced. See also AAQ39134-48
CC and AAR33982-91.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 278 AA;
XX
XX Query Match 100.0%; Score 20; DB 14; Length 278;
XX Best Local Similarity 100.0%; Pred. No. 2e-12;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 WHYPPKPCGIVPAKSVCGPV 20
XX |||||
XX DB 118 WHYPPKPCGIVPAKSVCGPV 137
XX
XX RESULT 13
XX AAR45330
XX ID AAR45330 standard; protein; 319 AA.
XX AC AAR45330;
XX
XX DT 28-JUN-1994 (first entry)
XX DE Anti-HCV antibody reactive protein #1.
XX KW Hepatitis C virus; HCV; envelope; region; anti-HCV; antibody; vaccine.
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX FT Peptide 39..69
XX FT /label= EP1
XX FT /note= "Claim 1"
XX
XX PN JP05310786-A.
XX PD 22-NOV-1993.
XX
XX PF 30-APR-1992; 92JP-0111846.
XX
XX PR 30-APR-1992; 92JP-0111846.
XX
XX PA (OLYU) OLYMPUS OPTICAL CO LTD.
XX
XX DR WPI; 1994-002315/01.
XX
XX PT Peptide of HCV envelope region - having high reactivity against
XX PT anti-hepatitis C virus antibody and having at least 6 amino acids
XX
XX PS Disclosure; Page 6-7; 11pp; Japanese.
XX
XX CC The sequences given in AAR45330-31 represent fragments derived from
XX CC the hepatitis C virus (HCV) envelope region that are highly reactive
XX CC against anti-HCV antibodies. These protein fragments may be used in
XX CC vaccines against HCV. They have a high reactivity against anti-HCV
XX CC antibodies regardless of the original HCV. They can be used to
XX CC detect HCV infection but not the onset of infection.
XX
XX Sequence 319 AA;

Query Match 100.0%; Score 20; DB 15; Length 319;
Best Local Similarity 100.0%; Pred. No. 2.2e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPCGIVPAKSVCGPV 20
DB 296 WHYPPKPCGIVPAKSVCGPV 315

XX RESULT 14
XX AAB68042
XX ID AAB68042 standard; protein; 363 AA.
XX AC AAB68042;
XX
XX DT 29-JUN-2001 (first entry)
XX DE Amino acid sequence of a Hepatitis C virus envelope E2 protein.
XX KW Eo protein; HCV; envelope protein; E2 protein; HCV infection;
XX KW HCV attachment.
XX
XX OS Hepatitis C virus.
XX PN WO200122984-A1.
XX
XX PD 05-APR-2001.
XX
XX PF 26-SEP-2000; 2000WO-US26395.
XX
XX PR 29-SEP-1999; 99US-0407430.
XX
XX PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
XX PI Worman HJ, Mamiya N;
XX
XX DR WPI; 2001-273486/28.
XX
XX PT Treating or preventing hepatitis C virus infection in a subject.
XX PT involves administering hepatitis C virus envelope protein E2 binding
XX agents -
XX
XX PS Claim 3; Fig 7; 46pp; English.
XX
XX CC The present sequence represents a Hepatitis C virus (HCV) envelope
XX CC E2 protein. Agents that bind to the HCV E2 protein (such as the human
XX CC Eo protein), and so inhibit the attachment of HCV onto cells
XX CC (especially liver cells), are used to treat HCV infections in mammals,
XX CC in particular humans. The specification also describes a method for
XX CC identifying a compound which can be used for treating or preventing
XX CC HCV in a subject and which can inhibit the attachment of HCV onto cells
XX CC by inhibiting the binding of HCV envelope E2 protein to a cellular
XX CC protein associated with HCV attachment and entry into cells. The method
XX CC comprises incubating the compound, HCV envelope E2 protein or its variant
XX CC and a cellular protein capable of specifically binding to the HCV E2
XX CC protein under suitable reaction conditions; determining the interactions
XX CC between HCV envelope E2 protein and cellular protein in the presence and
XX CC absence of the compound; and comparing the interaction to identify a
XX CC compound which can inhibit the attachment of HCV onto cells.
XX
XX SQ Sequence 363 AA;

XX Query Match 100.0%; Score 20; DB 22; Length 363;
XX Best Local Similarity 100.0%; Pred. No. 2.4e-12;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPCGIVPAKSVCGPV 20
DB 104 WHYPPKPCGIVPAKSVCGPV 123

XX RESULT 15

AAR34438
ID AAR34438 standard; Protein; 402 AA.
XX
AC AAR34438;
XX
XX 25-MAR-2003 (updated)
DT 09-AUG-1993 (first entry)
XX
XX Sequence of glycoprotein E2/NS1 in clone HCV1.
DE
XX Hepatitis C virus; envelope protein; glycoprotein; E2/NS1;
KW diagnostic reagent.
RW
XX Hepatitis C virus.
OS
XX EP537626-A1.
PN
XX
PD 21-APR-1993.
XX
XX 08-OCT-1992; 92EP-0117191.
PF
XX 08-OCT-1991; 91JP-0260824.
PR
XX (NAHE-) NAT INST OF HEALTH.
PA
XX Harada S, Honda Y, Miyamura T, Saito I;
PI
XX WPI; 1993-127516/16.
DR
XX N-PSDB; AAQ40330.
XX
PT Diagnostic reagent for hepatitis C virus - comprises second
PT envelope protein or first non-structural protein encoded by HCV
PT gene and has sugar chain
XX
PS Claim 2; Pages 30-32; 59pp; English.
XX
XX Glycoprotein E2/NS1 is derived from the second envelope protein or
CC first non-structural protein encoded by the genome of HCV. The
CC nucleic acid is extracted from the serum of the patient of hepatitis
CC C. The serum is pref. mixed with transfer RNA (tRNA) as a carrier
CC of virus RNA. As a technique of cloning cDNA from the nucleic acid,
CC it is preferred to use polymerase chain reaction method. In the
CC reaction, any commercially available random primers or synthesized
CC DNA having a base sequence similar to that of primer AS1 may be used
CC as a primer. Representative examples of sense primers include S1.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 402 AA;
Query Match 100.0%; Score 20; DB 14; Length 402;
Best Local Similarity 100.0%; Pred. No. 2.6e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WHYPPKPCGIVPAKSVCGPV 20
Db ||||||||||||||||
148 WHYPPKPCGIVPAKSVCGPV 167

Search completed: November 21, 2003, 20:58:04
Job time : 32.15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.6 Seconds
(without alignments)
79.832 Million cell updates/sec

Title: US-09-973-025-79

Perfect score: 20

Sequence: 1 WHYPPKCGIVPAKSVCGPV 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-08-612-973-79
2	20	100.0	20	3	US-08-927-597-79
3	20	100.0	20	4	US-08-635-886C-35
4	20	100.0	26	1	US-08-262-037-37
5	20	100.0	34	1	US-08-262-037-38
6	20	100.0	42	1	US-08-262-037-1
7	20	100.0	50	1	US-08-262-037-39
8	20	100.0	57	1	US-08-262-037-40
9	20	100.0	76	3	US-08-444-818-79
10	20	100.0	144	1	US-08-440-103-20
11	20	100.0	144	1	US-08-440-542-20
12	20	100.0	144	1	US-08-231-368-20
13	20	100.0	144	1	US-08-440-210-20
14	20	100.0	144	4	US-09-046-604-20
15	20	100.0	174	1	US-08-460-806-19
16	20	100.0	174	1	US-08-325-630-19
17	20	100.0	180	2	US-08-483-695-40
18	20	100.0	180	2	US-07-965-285-40
19	20	100.0	180	2	US-08-487-231-40
20	20	100.0	180	3	US-09-201-912-40
21	20	100.0	278	1	US-08-440-103-15
22	20	100.0	278	1	US-08-440-542-15
23	20	100.0	278	1	US-08-231-368-15
24	20	100.0	278	1	US-08-440-210-15
25	20	100.0	278	4	US-09-046-604-15
26	20	100.0	402	1	US-08-460-806-13
27	20	100.0	402	1	US-08-325-630-13

28	20	100.0	403	2	US-08-483-695-39	Sequence 39, Appl
29	20	100.0	403	2	US-07-965-285-39	Sequence 39, Appl
30	20	100.0	403	2	US-08-487-231-39	Sequence 39, Appl
31	20	100.0	403	3	US-09-201-912-39	Sequence 39, Appl
32	20	100.0	454	3	US-08-444-818-73	Sequence 73, Appl
33	20	100.0	463	1	US-07-748-292-9	Sequence 9, Appl
34	20	100.0	463	3	US-08-867-611-31	Sequence 31, Appl
35	20	100.0	463	5	PCT-US92-06965A-1	Sequence 1, Appl
36	20	100.0	480	I	US-08-440-103-14	Sequence 14, Appl
37	20	100.0	480	1	US-08-440-542-14	Sequence 14, Appl
38	20	100.0	480	1	US-08-231-368-14	Sequence 14, Appl
39	20	100.0	480	1	US-08-440-210-14	Sequence 14, Appl
40	20	100.0	480	4	US-09-046-604-14	Sequence 14, Appl
41	20	100.0	621	1	US-07-748-292-7	Sequence 7, Appl
42	20	100.0	622	3	US-08-867-611-34	Sequence 34, Appl
43	20	100.0	622	5	PCT-US92-06965A-4	Sequence 4, Appl
44	20	100.0	663	3	US-08-824-057-3	Sequence 3, Appl
45	20	100.0	663	4	US-09-415-582-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-612-973-79
; Sequence 79, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P. C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-612-973-79

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKCGIVPAKSVCGPV 20
|||||

Db 1 WHYPKPCGIVPAKSVCGPV 20

RESULT 2

US-08-927-597-79
; Sequence 79, Application US/08927597
; Patent No. 6245503

GENERAL INFORMATION:

; APPLICANT: MAERTENS, GEERT

; APPLICANT: BOSMAN, FONS

; APPLICANT: DE MARTYNOFF, GUY

; APPLICANT: BUYSE, MARIE-ANGE

; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

; NUMBER OF SEQUENCES: 111

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: NIXON & VANDERHYE P.C.

; STREET: 1100 NORTH GLEBE ROAD

; CITY: ARLINGTON

; STATE: VIRGINIA

; COUNTRY: U.S.A.

; ZIP: 22201-4714

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/927,597

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/612,973

; FILING DATE: 11-MAR-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: BYRNE, THOMAS E.

; REGISTRATION NUMBER: 32,205

; REFERENCE/DOCKET NUMBER: 1487-10

; TELEPHONE: (703) 816-4000

; TELEFAX: (703) 816-4100

; INFORMATION FOR SEQ ID NO: 79:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 20 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-927-597-79

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGPV 20

Db 1 WHYPKPCGIVPAKSVCGPV 20

RESULT 3

US-08-635-886C-35
; Sequence 35, Application US/08635886C
; Patent No. 6555114

GENERAL INFORMATION:

; APPLICANT: LEROUX-ROELS, Geert

; APPLICANT: DELEYS, Robert

; APPLICANT: MAERTENS, Geert

; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C

; TITLE OF INVENTION: VIRUS

; FILE REFERENCE: 2752-18

; CURRENT APPLICATION NUMBER: US/08/635,886C

; CURRENT FILING DATE: 1996-04-25

; PRIOR APPLICATION NUMBER: PCT/EP94/03555

; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6

; NUMBER OF SEQ ID NOS: 286

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 35

; LENGTH: 20

; TYPE: PRT

; ORGANISM: hepatitis C virus

US-08-635-886C-35

Query Match 100.0%; Score 20; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 2e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGPV 20

Db 1 WHYPKPCGIVPAKSVCGPV 20

RESULT 4

US-08-262-037-37

; Sequence 37, Application US/08262037

; Patent No. 5747239

; GENERAL INFORMATION:

; APPLICANT: Chang Yi Wang and Barbara Hosen

; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR

; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV

; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES

; NUMBER OF SEQUENCES: 136

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNEGAN

; STREET: 345 PARK AVE.

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10154

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: WORDPERFECT 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/262,037

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/719,819

; FILING DATE: 24-June-1991

; APPLICATION NUMBER: 07/667,275

; FILING DATE: 11-Mar-1991

; APPLICATION NUMBER: 07/651,735

; FILING DATE: 07-Feb-1991

; APPLICATION NUMBER: 07/558,799

; FILING DATE: 26-July-1990

; APPLICATION NUMBER: 07/510,153

; FILING DATE: 16-April-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Maria C. H. Lin

; REGISTRATION NUMBER: 29,323

; REFERENCE/DOCKET NUMBER: 1151-4043 US3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-758-4800

; TELEFAX: (212) 751-6849

; TELEX: 421792

; INFORMATION FOR SEQ ID NO: 37:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 26 amino acids

; TYPE: Amino acid

; STRANDEDNESS:

; TOPOLOGY: Unknown

US-08-262-037-37

```
Query Match      100.0%; Score 20; DB 1; Length 26;
Best Local Similarity 100.0%; Pred. No. 2.4e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPKPCGIVPAKSVCGVP 20
Db 5 WHYPKPCGIVPAKSVCGVP 24
|||||

RESULT 5
US-08-262-037-38
; Sequence 38, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide
; HYPOTHETICAL: No
; ANTI-SENSE: No
; US-08-262-037-1

Query Match      100.0%; Score 20; DB 1; Length 34;
Best Local Similarity 100.0%; Pred. No. 3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPKPCGIVPAKSVCGVP 20
Db 13 WHYPKPCGIVPAKSVCGVP 32
|||||

RESULT 6
US-08-262-037-1
; Sequence 1, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide
; HYPOTHETICAL: No
; ANTI-SENSE: No
; US-08-262-037-1

Query Match      100.0%; Score 20; DB 1; Length 42;
Best Local Similarity 100.0%; Pred. No. 3.6e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPKPCGIVPAKSVCGVP 20
Db 21 WHYPKPCGIVPAKSVCGVP 40
|||||

RESULT 7
US-08-262-037-39
; Sequence 39, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
```

Query Match 100.0%; Score 20; DB 1; Length 50;
Best Local Similarity 100.0%; Pred. No. 4.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPPKCGIVPAKSVCGPV 20
 |||||
Db 29 WHYPPKCGIVPAKSVCGPV 48
 |||||

RESULT 8
US-08-262-037-40
; Sequence 40, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

: Sequence 79, Application US/08444818
: Patent No. 6150087
: GENERAL INFORMATION:
: APPLICANT: Chien, David Y.
: APPLICANT: Chien, William J.
: TITLE OF INVENTION: NANEV Diagnostics and Vaccines
: NUMBER OF SEQUENCES: 777
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Chiron Corporation
: STREET: 4560 Horton Street
: CITY: Emeryville
: STATE: CA
: COUNTRY: USA
: ZIP: 94608-2916
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/444,818
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/403,590
: FILING DATE: 14-MAR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Harbin, Alisa A.
: REGISTRATION NUMBER: 33,995
: REFERENCE/DOCKET NUMBER: 0110.002

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (508)359-3876
TELEFAX: (508)359-3885
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 76 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-444-818-79

Query Match 100.0%; Score 20; DB 3; Length 76;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPKPCGIVPAKSVCGVPV 20
DB 47 WHYPKPCGIVPAKSVCGVPV 66

RESULT 10
US-08-440-103-20
Sequence 20, Application US/08440103
Patent No. 5670152
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICANT: Weiner, Amy J.
FILING DATE: 13-SEP-1991
CLASSIFICATION: 424
PRIOR APPLICATION NUMBER: US/08/440,103
APPLICATION NUMBER: US/08/231,368
FILING DATE:
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-103-20

Query Match 100.0%; Score 20; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 9.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPKPCGIVPAKSVCGVPV 20
DB 118 WHYPKPCGIVPAKSVCGVPV 137

RESULT 11
US-08-440-542-20
Sequence 20, Application US/08440542
Patent No. 5670153
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICANT: Weiner, Amy J.
FILING DATE: 12-MAY-1995
CLASSIFICATION: 424
PRIOR APPLICATION NUMBER: US/08/440,542
APPLICATION NUMBER: US/08/231,368
FILING DATE:
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-542-20

Query Match 100.0%; Score 20; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 9.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPKPCGIVPAKSVCGVPV 20
DB 118 WHYPKPCGIVPAKSVCGVPV 137

RESULT 12
US-08-231-368-20
Sequence 20, Application US/08231368
Patent No. 5756312
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608

Query Match 100.0%; Score 20; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 9.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPKPCGIVPAKSVCGVPV 20
DB 118 WHYPKPCGIVPAKSVCGVPV 137

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-231-368-20

Query Match 100.0%; Score 20; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 9.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGPV 20
Db 118 WHYPKPCGIVPAKSVCGPV 137

RESULT 13
US-08-440-210-20
; Sequence 20, Application US/08440210
; Patent No. 5766845
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,210
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELEPHONE: (510) 601-2708
```

```
;
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-440-210-20

Query Match 100.0%; Score 20; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 9.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGPV 20
Db 118 WHYPKPCGIVPAKSVCGPV 137

RESULT 14
US-09-046-604-20
; Sequence 20, Application US/09046604
; Patent No. 6303292
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/046,604
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-09-046-604-20

Query Match 100.0%; Score 20; DB 4; Length 144;
Best Local Similarity 100.0%; Pred. No. 9.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGPV 20
Db 118 WHYPKPCGIVPAKSVCGPV 137
```


RESULT 15
US-08-460-806-19
; Sequence 19, Application US/08460806
; Patent No. 5747241
; GENERAL INFORMATION:
; APPLICANT: MIYAMURA, TATSUO
; APPLICANT: SAITO, IZUMU
; APPLICANT: HARADA, SHIZUKO
; APPLICANT: HONDA, YOSHIKAZU
; TITLE OF INVENTION: DIAGNOSTIC REAGENT FOR HEPATITIS C
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,806
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,630
; FILING DATE: 19-OCT-1994
; APPLICATION NUMBER: US 07/956,993
; FILING DATE: 06-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5747241man P.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 4667-001-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 174 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-460-806-19

Query Match 100.0%; Score 20; DB 1; Length 174;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 WHYPPKPGIVPAKSVCGPV 20
Db 148 WHYPPKPGIVPAKSVCGPV 167

Search completed: November 21, 2003, 21:15:15
Job time : 10.6 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time 19.5 Seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-79

Perfect score: 20

Sequence: 1 WHYPKPGCIVPAKSVCGPV 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	10	US-09-973-025-79
2	20	100.0	20	11	US-09-899-303-79
3	20	100.0	20	11	US-09-995-808-79
4	20	100.0	20	11	US-09-995-860-79
5	20	100.0	20	12	US-09-995-791-79
6	20	100.0	254	10	US-09-407-430-3
7	20	100.0	363	10	US-09-407-430-2
8	20	100.0	637	12	US-10-187-257-4
9	20	100.0	637	12	US-10-265-083-2
10	20	100.0	2894	10	US-09-941-611-23
11	20	100.0	2894	15	US-10-044-995-23
12	20	100.0	3011	9	US-09-916-359-2
13	20	100.0	3011	16	US-10-232-643-6
14	14	70.0	26	10	US-09-921-397-7
15	14	70.0	26	10	US-09-921-397-35

16	14	70.0	29	10	US-09-921-397-24	Sequence 24, Appl
17	14	70.0	33	10	US-09-921-397-14	Sequence 14, Appl
18	14	70.0	37	10	US-09-921-397-36	Sequence 36, Appl
19	14	70.0	96	10	US-09-921-397-82	Sequence 82, Appl
20	14	70.0	176	10	US-09-921-397-81	Sequence 81, Appl
21	14	70.0	250	10	US-09-952-572-8	Sequence 8, Appl
22	14	70.0	350	10	US-09-929-955-4	Sequence 4, Appl
23	14	70.0	350	14	US-10-104-966-4	Sequence 4, Appl
24	14	70.0	363	12	US-10-128-587A-97	Sequence 97, Appl
25	14	70.0	363	15	US-10-128-590-97	Sequence 97, Appl
26	14	70.0	3011	9	US-09-742-659-4	Sequence 4, Appl
27	14	70.0	3011	10	US-09-238-076-20	Sequence 20, Appl
28	14	70.0	3011	10	US-09-952-572-9	Sequence 9, Appl
29	14	70.0	3011	10	US-09-929-955-1	Sequence 1, Appl
30	14	70.0	3011	10	US-09-747-419-20	Sequence 20, Appl
31	14	70.0	3011	11	US-09-891-894-3	Sequence 3, Appl
32	14	70.0	3011	11	US-09-995-937-20	Sequence 20, Appl
33	14	70.0	3011	11	US-09-917-563-20	Sequence 20, Appl
34	14	70.0	3011	12	US-10-184-150-3	Sequence 3, Appl
35	14	70.0	3011	14	US-10-104-966-1	Sequence 1, Appl
36	14	70.0	3011	15	US-10-259-275-20	Sequence 20, Appl
37	14	70.0	3012	10	US-09-238-076-2	Sequence 2, Appl
38	14	70.0	3012	11	US-09-995-937-2	Sequence 2, Appl
39	14	70.0	3012	11	US-09-917-563-2	Sequence 22, Appl
40	12	60.0	43	10	US-09-921-397-22	Sequence 78, Appl
41	8	40.0	20	10	US-09-973-025-78	Sequence 80, Appl
42	8	40.0	20	10	US-09-973-025-80	Sequence 78, Appl
43	8	40.0	20	11	US-09-899-303-78	Sequence 80, Appl
44	8	40.0	20	11	US-09-899-303-80	Sequence 78, Appl
45	8	40.0	20	11	US-09-995-808-78	Sequence 78, Appl

ALIGNMENTS

RESULT 1

US-09-973-025-79
; Sequence 79, Application US/09973025
; Publication No. US20020182706A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; o COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; *APPLICATION NUMBER: US/09/973,025
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-09-973-025-79

Query Match 100.0%; Score 20; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPKPCGIVPAKSVCGPV 20
|||||
DB 1 WHYPKPCGIVPAKSVCGPV 20

RESULT 2
US-09-899-303-79
; Sequence 79, Application US/09899303
; Publication No. US20030036110A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GREERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-09-899-303-79

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WHYPKPCGIVPAKSVCGPV 20
|||||

Db 1 WHYPKPCGIVPAKSVCGPV 20

RESULT 3
US-09-995-808-79
; Sequence 79, Application US/09995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 79
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-79

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPKPCGIVPAKSVCGPV 20
|||||
DB 1 WHYPKPCGIVPAKSVCGPV 20

RESULT 4
US-09-995-860-79
; Sequence 79, Application US/09995860
; Publication No. US20030118603A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; therapeutic use.
; FILE REFERENCE: 2551-89
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 79
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-79

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPKPCGIVPAKSVCGPV 20
|||||
DB 1 WHYPKPCGIVPAKSVCGPV 20

RESULT 5
US-09-995-791-79
; Sequence 79, Application US/09995791
; Publication No. US20030147918A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; therapeutic use.
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 79
US-09-995-791-79

```
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-79

Query Match      100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGVP 20
Db 1 WHYPKPCGIVPAKSVCGVP 20

RESULT 6
US-09-407-430-3
; Sequence 3, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; TITLE OF INVENTION: HEPATITIS C VIRUS INFECTION
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-407-430-3

Query Match      100.0%; Score 20; DB 10; Length 254;
Best Local Similarity 100.0%; Pred. No. 4.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGVP 20
Db 82 WHYPKPCGIVPAKSVCGVP 101

RESULT 7
US-09-407-430-2
; Sequence 2, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; TITLE OF INVENTION: HEPATITIS C VIRUS INFECTION
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-09-407-430-2

Query Match      100.0%; Score 20; DB 10; Length 363;
Best Local Similarity 100.0%; Pred. No. 6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGVP 20
Db 104 WHYPKPCGIVPAKSVCGVP 123

RESULT 8
US-10-187-257-4
; Sequence 4, Application US/10187257
; Publication No. US20030138458A1
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: O'HAGAN, Derek
; TITLE OF INVENTION: HCV E1E2 VACCINE COMPOSITIONS
; FILE REFERENCE: 2302-17206
; CURRENT APPLICATION NUMBER: US/10/187,257
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
; OTHER INFORMATION: amino acid
US-10-187-257-4

Query Match      100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 9.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGVP 20
Db 315 WHYPKPCGIVPAKSVCGVP 334

RESULT 9
US-10-265-083-2
; Sequence 2, Application US/10265083
; Publication No. US20030170273A1
; GENERAL INFORMATION:
; APPLICANT: O'HAGAN, Derek
; APPLICANT: VALIANTE, Nicholas
; TITLE OF INVENTION: ADJUVANT COMPOSITIONS
; FILE REFERENCE: 2302-18357.30
; CURRENT APPLICATION NUMBER: US/10/265,083
; CURRENT FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7
; OTHER INFORMATION: region
US-10-265-083-2

Query Match      100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 9.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGVP 20
Db 315 WHYPKPCGIVPAKSVCGVP 334

RESULT 10
US-09-941-611-23
; Sequence 23, Application US/09941611
; Patent No. US20020106640A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSWUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; ANTIBODIES TO HEPATITIS C VIRUS
```

NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/941,611
FILING DATE: 30-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: 1995-02-21
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-941-611-23
Query Match 100.0%; Score 20; DB 10; Length 2894;
Best Local Similarity 100.0%; Pred. No. 3.3e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WHYPPKCGIVPAKSVCGPV 20
DB 487 WHYPPKCGIVPAKSVCGPV 506
RESULT 11
US-10-044-995-23
Sequence 23, Application US/10044995
Publication No. US20030049685A1
GENERAL INFORMATION:
APPLICANT: DELEYS, ROBERT J
POLLET, DIRK
MAERTENS, GEERT
VAN HEUVERSWUN, HUGO
TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
ANTIBODIES TO HEPATITIS C VIRUS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/044,995
FILING DATE: 15-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: <Unknown>
APPLICATION NUMBER: US 07/920,286
FILING DATE: 14-OCT-1992
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-044-995-23
Query Match 100.0%; Score 20; DB 15; Length 2894;
Best Local Similarity 100.0%; Pred. No. 3.3e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WHYPPKCGIVPAKSVCGPV 20
DB 487 WHYPPKCGIVPAKSVCGPV 506
RESULT 12
US-09-916-359-2
Sequence 2, Application US/09916359
Patent No. US20020034734A1
GENERAL INFORMATION:
APPLICANT: Veronique Barban
TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
TREATING C HEPATITIS
FILE REFERENCE: PMC97-03A
CURRENT APPLICATION NUMBER: US/09/916,359
CURRENT FILING DATE: 2001-07-26
PRIOR APPLICATION NUMBER: 09/388,874
PRIOR FILING DATE: 1999-09-02
PRIOR APPLICATION NUMBER: 97/02,887
PRIOR FILING DATE: 1997-03-06
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 3011
TYPE: PPT
ORGANISM: Virus
US-09-916-359-2
Query Match 100.0%; Score 20; DB 9; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WHYPPKCGIVPAKSVCGPV 20
DB 487 WHYPPKCGIVPAKSVCGPV 506

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```
;
;
; NAME/KEY: Duplication
; LOCATION: 2690
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Arg or Gly"
;
; FEATURE:
;
; NAME/KEY: Duplication
; LOCATION: 2921
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Arg or Gly"
;
; FEATURE:
;
; NAME/KEY: Duplication
; LOCATION: 2996
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Leu or Pro"
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-232-643-6

Query Match          100.0%; Score 20; DB 16; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGPV 20
Db 487 WHYPKPCGIVPAKSVCGPV 506

RESULT 14
US-09-921-397-7
; Sequence 7, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-7

Query Match          70.0%; Score 14; DB 10; Length 26;
Best Local Similarity 100.0%; Pred. No. 9.3e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 PCGIVPAKSVCGPV 20
Db 4 PCGIVPAKSVCGPV 17

RESULT 15
US-09-921-397-35
; Sequence 35, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
```

```
;
;
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-35

Query Match          70.0%; Score 14; DB 10; Length 26;
Best Local Similarity 100.0%; Pred. No. 9.3e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 PCGIVPAKSVCGPV 20
Db 5 PCGIVPAKSVCGPV 18

Search completed: November 21, 2003, 22:19:39
Job time : 19.5 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 163.1 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-79

Perfect score: 20

Sequence: 1 WHYPKPGCIVPAKSVCGPV 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents AA Main:*

- 1: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/1/paa/US097A_COMB.pep.*
- 22: /cgn2_6/ptodata/1/paa/US097B_COMB.pep.*
- 23: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
- 24: /cgn2_6/ptodata/1/paa/US099A_COMB.pep.*
- 25: /cgn2_6/ptodata/1/paa/US099B_COMB.pep.*
- 26: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
- 27: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
- 28: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
- 29: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
- 30: /cgn2_6/ptodata/1/paa/US104_COMB.pep.*
- 31: /cgn2_6/ptodata/1/paa/US106_COMB.pep.*
- 32: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of: results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	13	US-08-974-685-35
					Sequence 35, Appl

2	20	100.0	20	13	US-08-974-690-35	Sequence 35, Appl
3	20	100.0	20	13	US-08-974-690A-35	Sequence 35, Appl
4	20	100.0	20	13	US-08-974-690B-35	Sequence 35, Appl
5	20	100.0	20	13	US-08-974-690C-35	Sequence 35, Appl
6	20	100.0	20	23	US-09-899-303-79	Sequence 79, Appl
7	20	100.0	20	23	US-09-899-303A-79	Sequence 79, Appl
8	20	100.0	20	25	US-09-973-025-79	Sequence 79, Appl
9	20	100.0	20	25	US-09-973-025-79	Sequence 79, Appl
10	20	100.0	20	25	US-09-995-808-79	Sequence 79, Appl
11	20	100.0	20	25	US-09-995-808-79	Sequence 79, Appl
12	20	100.0	20	26	US-10-020-510-79	Sequence 79, Appl
13	20	100.0	20	29	US-10-321-798-79	Sequence 79, Appl
14	20	100.0	26	8	US-08-475-482-37	Sequence 37, Appl
15	20	100.0	26	8	US-08-477-072-37	Sequence 37, Appl
16	20	100.0	26	8	US-08-477-582-37	Sequence 37, Appl
17	20	100.0	26	8	US-08-480-253-37	Sequence 38, Appl
18	20	100.0	34	8	US-08-475-482-38	Sequence 38, Appl
19	20	100.0	34	8	US-08-477-072-38	Sequence 38, Appl
20	20	100.0	34	8	US-08-477-582-38	Sequence 38, Appl
21	20	100.0	34	8	US-08-480-253-38	Sequence 38, Appl
22	20	100.0	42	3	US-07-667-275A-1	Sequence 1, Appl
23	20	100.0	42	8	US-08-475-482-1	Sequence 1, Appl
24	20	100.0	42	8	US-08-477-072-1	Sequence 1, Appl
25	20	100.0	42	8	US-08-477-582-1	Sequence 1, Appl
26	20	100.0	44	1	PCT-US94-12268-10	Sequence 10, Appl
27	20	100.0	44	5	US-08-143-412A-10	Sequence 10, Appl
28	20	100.0	44	8	US-08-479-690-10	Sequence 10, Appl
29	20	100.0	50	8	US-08-475-482-39	Sequence 39, Appl
30	20	100.0	50	8	US-08-477-072-39	Sequence 39, Appl
31	20	100.0	50	8	US-08-477-582-39	Sequence 39, Appl
32	20	100.0	50	8	US-08-480-253-39	Sequence 40, Appl
33	20	100.0	57	8	US-08-475-482-40	Sequence 40, Appl
34	20	100.0	57	8	US-08-477-072-40	Sequence 40, Appl
35	20	100.0	57	8	US-08-477-582-40	Sequence 40, Appl
36	20	100.0	57	8	US-08-477-072-40	Sequence 40, Appl
37	20	100.0	57	8	US-08-480-253-40	Sequence 40, Appl
38	20	100.0	76	8	US-08-403-590B-79	Sequence 79, Appl
39	20	100.0	76	8	US-08-444-112-79	Sequence 79, Appl
40	20	100.0	122	8	US-08-436-966-5	Sequence 5, Appl
41	20	100.0	144	8	US-08-471-498-20	Sequence 20, Appl
42	20	100.0	254	1	PCT-US00-26395-3	Sequence 3, Appl
43	20	100.0	254	18	US-09-407-430-3	Sequence 3, Appl
44	20	100.0	278	8	US-08-471-498-15	Sequence 15, Appl
45	20	100.0	333	30	US-10-445-724-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1

US-08-974-685-35

Sequence 35, Application US/08974685

GENERAL INFORMATION:

APPLICANT: LEROUX-ROELS, GEERT

DELEYS, ROBERT

MAERTENS, GEERT

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF

HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 181

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHVE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,685
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 35:

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGPV 20
Db 1 WHYPKPCGIVPAKSVCGPV 20

RESULT 2

US-08-974-690-35
Sequence 35, Application US/08974690A

GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 166

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,690
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/635,886
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-11

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 35:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-974-690-35

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGPV 20
Db 1 WHYPKPCGIVPAKSVCGPV 20

RESULT 3

US-08-974-690A-35

Sequence 35, Application US/08974690A

GENERAL INFORMATION:

APPLICANT: LEROUX-ROELS, GEERT

APPLICANT: DELEYS, ROBERT

MAERTENS, GEERT

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF

HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 177

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,690A

FILING DATE: 19-Nov-1997

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: SADOFF, B. J.

REGISTRATION NUMBER: 36,663

REFERENCE/DOCKET NUMBER: 2752-20

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 35:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 35:

US-08-974-690A-35

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGPV 20
Db 1 WHYPKPCGIVPAKSVCGPV 20

RESULT 4

US-08-974-690B-35

Sequence 35, Application US/08974690B

GENERAL INFORMATION:

APPLICANT: LEROUX-ROELS, GEERT

APPLICANT: DELEYS, ROBERT

MAERTENS, GEERT

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF

HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690B
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US-08-974-690B-35

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPPKPGIVPAKSVCGPV 20
Db 1 WHYPPKPGIVPAKSVCGPV 20

RESULT 5
US-08-974-690C-35
Sequence 35, Application US/08974690C
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELEYS, Robert
APPLICANT: MAERTENS, Geert
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
FILE REFERENCE: 2551-94
CURRENT APPLICATION NUMBER: US/08/974,690C
CURRENT FILING DATE: 1997-11-19
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28
PRIOR APPLICATION NUMBER: EP 93402718.6
PRIOR FILING DATE: 1993-11-04
NUMBER OF SEQ ID NOS: 286
SOFTWARE: PatentIn version 3.1
SEQ ID NO 35
LENGTH: 20
TYPE: PRT
ORGANISM: hepatitis C virus
US-08-974-690C-35

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPPKPGIVPAKSVCGPV 20

Db 1 WHYPPKPGIVPAKSVCGPV 20

RESULT 6
US-09-899-303-79
Sequence 79, Application US/09899303
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-09-899-303-79

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPPKPGIVPAKSVCGPV 20
Db 1 WHYPPKPGIVPAKSVCGPV 20

RESULT 7
US-09-899-303A-79
Sequence 79, Application US/09899303A
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.

STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303A
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-09-899-303A-79
Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WHYPKPCGIVPAKSVCGPV 20
DB 1 WHYPKPCGIVPAKSVCGPV 20
RESULT 8
US-09-973-025-79
Sequence 79, Application US/09973025
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSER: NIXON & VANDERHUYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/973,025
FILING DATE: 10-OCT-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-09-973-025-79
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WHYPKPCGIVPAKSVCGPV 20
DB 1 WHYPKPCGIVPAKSVCGPV 20
RESULT 9
US-09-995-791-79
Sequence 79, Application US/09995791
GENERAL INFORMATION:
APPLICANT: Innogenetics N.V.
TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
therapeutic use.
FILE REFERENCE: 2551-68
CURRENT APPLICATION NUMBER: US/09/995,791
CURRENT FILING DATE: 2001-11-29
NUMBER OF SEQ ID NOS: 122
SOFTWARE: Patentin 3.1
SEQ ID NO 79
LENGTH: 20
TYPE: PRT
ORGANISM: Hepatitis C virus
US-09-995-791-79
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WHYPKPCGIVPAKSVCGPV 20
DB 1 WHYPKPCGIVPAKSVCGPV 20
RESULT 10
US-09-995-808-79
Sequence 79, Application US/09995808
GENERAL INFORMATION:
APPLICANT: Innogenetics N.V.
TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
therapeutic use.
FILE REFERENCE: 2551-70
CURRENT APPLICATION NUMBER: US/09/995,808
CURRENT FILING DATE: 2001-11-29
NUMBER OF SEQ ID NOS: 122
SOFTWARE: Patentin 3.1
SEQ ID NO 79
LENGTH: 20
TYPE: PRT
ORGANISM: Hepatitis C virus
US-09-995-808-79
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.1e-14;

```
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WHYPKPCGIVPAKSVCGPV 20
Db 1 WHYPKPCGIVPAKSVCGPV 20

RESULT 11
US-09-995-860-79
; Sequence 79, Application US/09995860
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 79
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-79

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WHYPKPCGIVPAKSVCGPV 20
Db 1 WHYPKPCGIVPAKSVCGPV 20

RESULT 12
US-10-020-510-79
; Sequence 79, Application US/10020510
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-72
; CURRENT APPLICATION NUMBER: US/10/020,510
; CURRENT FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 79
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-020-510-79

Query Match 100.0%; Score 20; DB 26; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WHYPKPCGIVPAKSVCGPV 20
Db 1 WHYPKPCGIVPAKSVCGPV 20

RESULT 13
US-10-321-798-79
; Sequence 79, Application US/10321798
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-93
; CURRENT APPLICATION NUMBER: US/10/321,798
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: 60/418,358
; PRIOR FILING DATE: 2002-10-16
```

```
; PRIOR APPLICATION NUMBER: 10/020,510
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 79
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-321-798-79

Query Match 100.0%; Score 20; DB 29; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WHYPKPCGIVPAKSVCGPV 20
Db 1 WHYPKPCGIVPAKSVCGPV 20

RESULT 14
US-08-475-482-37
; Sequence 37, Application US/08475482
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION NUMBER: US/08/475,482
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 17-June-1994
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-475-482-37
```

Job time : 163.1 secs

Query Match 100.0%; Score 20; DB 8; Length 26;
Best Local Similarity 100.0%; Pred. No. 6.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPGIIVPAKSVCGPV 20
DB 5 WHYPPKPGIIVPAKSVCGPV 24

RESULT 15
US-08-477-072-37
; Sequence 37, Application US/08477072
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,072
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 17-June-1994
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-477-072-37

Query Match 100.0%; Score 20; DB 8; Length 26;
Best Local Similarity 100.0%; Pred. No. 6.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPGIIVPAKSVCGPV 20
DB 5 WHYPPKPGIIVPAKSVCGPV 24

Search completed: November 21, 2003, 22:09:51

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.55 Seconds
(without alignments)
117,536 Million cell updates/sec

Title: US-09-973-025-79

Perfect score: 20

Sequence: 1 WHYPKPGIVPAKSVCGPV 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents AA_New.*

- 1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	6	US-10-651-165-35
2	20	100.0	30	6	US-10-296-734-472
3	20	100.0	333	1	PCT-US03-19834-3
4	20	100.0	333	6	US-10-655-562-4
5	20	100.0	637	1	PCT-US03-33610-4
6	20	100.0	2010	6	US-10-296-734-814
7	20	100.0	3011	1	PCT-US03-19834-2
8	20	100.0	3011	6	US-10-296-734-406
9	20	100.0	5985	6	US-10-296-734-810
10	13	65.0	30	6	US-10-296-734-474
11	8	40.0	20	6	US-10-651-165-34
12	8	40.0	20	6	US-10-651-165-36
13	7	35.0	30	6	US-10-296-734-470
14	7	35.0	34	6	US-10-685-435-21
15	7	35.0	347	6	US-10-664-391-9
16	7	35.0	539	6	US-10-664-391-11
17	6	30.0	285	6	US-10-679-063-2355
18	6	30.0	342	6	US-10-425-114A-47952
19	6	30.0	867	5	US-09-581-286A-455
20	6	30.0	875	5	US-09-581-286A-328
21	6	30.0	950	6	US-10-679-063-12565
22	6	30.0	952	6	US-10-679-063-12566
23	6	30.0	1448	6	US-10-679-063-23423
24	6	30.0	2280	1	PCT-US03-20322-211
25	5	25.0	31	6	US-10-685-435-22
26	5	25.0	71	1	PCT-US03-30720-2535

27	5	25.0	105	6	US-10-425-114A-61830	Sequence 61830, A
28	5	25.0	106	6	US-10-425-114A-63148	Sequence 63148, A
29	5	25.0	110	6	US-10-425-114A-47730	Sequence 47730, A
30	5	25.0	122	1	PCT-US03-30870-52	Sequence 52, Appl
31	5	25.0	140	6	US-10-679-063-20560	Sequence 20560, A
32	5	25.0	160	6	US-10-425-114A-38672	Sequence 38672, A
33	5	25.0	167	6	US-10-679-063-12873	Sequence 12873, A
34	5	25.0	168	6	US-10-679-063-20121	Sequence 20121, A
35	5	25.0	176	5	US-09-638-693-158	Sequence 158, App
36	5	25.0	176	5	US-09-873-224A-158	Sequence 158, App
37	5	25.0	208	6	US-10-425-114A-68403	Sequence 68403, A
38	5	25.0	216	1	PCT-US03-33947-838	Sequence 838, App
39	5	25.0	217	6	US-10-425-114A-52869	Sequence 52869, A
40	5	25.0	221	6	US-10-425-114A-47997	Sequence 47997, A
41	5	25.0	223	5	US-09-674-546A-524	Sequence 524, App
42	5	25.0	228	6	US-10-679-063-14948	Sequence 14948, A
43	5	25.0	228	6	US-10-679-063-19541	Sequence 19541, A
44	5	25.0	236	6	US-10-679-063-18131	Sequence 18131, A
45	5	25.0	236	6	US-10-679-063-18132	Sequence 18132, A

ALIGNMENTS

RESULT 1
US-10-651-165-35
; Sequence 35, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELSIS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-35

Query Match 100.0%; Score 20; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPGIVPAKSVCGPV 20
Db 1 WHYPKPGIVPAKSVCGPV 20

RESULT 2
US-10-296-734-472
; Sequence 472, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 472
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 33
US-10-296-734-472

Query Match 100.0%; Score 20; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGVP 20
Db 9 WHYPKPCGIVPAKSVCGVP 28
|||||

RESULT 3
PCT-US03-19834-3
; Sequence 3, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; TITLE OF INVENTION: METHODS FOR TREATING HYPERLIPIDEMIA
; FILE REFERENCE: IOWA:045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-3

Query Match 100.0%; Score 20; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGVP 20
Db 105 WHYPKPCGIVPAKSVCGVP 124
|||||

RESULT 4
US-10-655-562-4
; Sequence 4, Application US/10655562
; GENERAL INFORMATION:
; APPLICANT: BISHOP-HURLEY, SHARON L.
; APPLICANT: SCHMIDT, FRANCIS J.
; APPLICANT: SMITH, ARNOLD L.
; TITLE OF INVENTION: PHAGE-DISPLAY PEPTIDES AS NOVEL ANTIMICROBIAL AGENTS
; TITLE OF INVENTION: AGAINST HAEMOPHILUS INFLUENZAE
; FILE REFERENCE: UVM0:022US
; CURRENT APPLICATION NUMBER: US/10/655,562
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: 60,409,909
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-655-562-4

Query Match 100.0%; Score 20; DB 6; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGVP 20
Db 105 WHYPKPCGIVPAKSVCGVP 124
|||||

RESULT 5
PCT-US03-33610-4
; Sequence 4, Application PC/TUS0333610
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: SELBY, Mark
; APPLICANT: PALIARD, Xavier
; TITLE OF INVENTION: ACTIVATION OF HCV-SPECIFIC T-CELLS
; FILE REFERENCE: 2300-1612.60
; CURRENT APPLICATION NUMBER: PCT/US03/33610
; CURRENT FILING DATE: 2003-10-24
; PRIOR APPLICATION NUMBER: 10/281,341
; PRIOR FILING DATE: 2002-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
PCT-US03-33610-4

Query Match 100.0%; Score 20; DB 1; Length 637;
Best Local Similarity 100.0%; Pred. No. 2.4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGVP 20
Db 315 WHYPKPCGIVPAKSVCGVP 334
|||||

RESULT 6
US-10-296-734-814
; Sequence 814, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 814
; LENGTH: 2010
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette B
US-10-296-734-814

Query Match 100.0%; Score 20; DB 6; Length 2010;
Best Local Similarity 100.0%; Pred. No. 5.1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPKPCGIVPAKSVCGVP 20
Db 969 WHYPKPCGIVPAKSVCGVP 988
|||||

Query Match	Best Local Similarity	Score	DB	Length	Indels	Gaps
Query Match	Best Local Similarity	100.0%;	Score 20;	DB 1;	Length 3011;	
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	WHYPPKPGGIVPAKSVCGPV 20				
DB	487	WHYPPKPGGIVPAKSVCGPV 506				
RESULT 8						
US-10-296-734-406						
Sequence 406, Application US/10296734						
GENERAL INFORMATION:						
APPLICANT: Thompson, Scott A						
APPLICANT: Ramshaw, Ian A						
TITLE OF INVENTION: Synthetic molecules and uses therefor						
FILE REFERENCE: Savine						
CURRENT APPLICATION NUMBER: US/10/296,734						
PRIOR FILING DATE: 2003-08-04						
PRIOR APPLICATION NUMBER: AU PQ7761/00						
PRIOR FILING DATE: 2000-05-26						
NUMBER OF SEQ ID NOS: 1507						
SOFTWARE: PatentIn version 3.2						
SEQ ID NO 406						
LENGTH: 3011						
TYPE: PRT						
ORGANISM: Artificial						
FEATURE:						
OTHER INFORMATION: HepC 1a consensus polyprotein						
US-10-296-734-406						
Query Match	Best Local Similarity	100.0%;	Score 20;	DB 6;	Length 3011;	
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	WHYPPKPGGIVPAKSVCGPV 20				
DB	487	WHYPPKPGGIVPAKSVCGPV 506				
RESULT 9						
US-10-296-734-810						
Sequence 810, Application US/10296734						
GENERAL INFORMATION:						
APPLICANT: Thompson, Scott A						
APPLICANT: Ramshaw, Ian A						
TITLE OF INVENTION: Synthetic molecules and uses therefor						
FILE REFERENCE: Savine						
CURRENT APPLICATION NUMBER: US/10/296,734						
PRIOR FILING DATE: 2003-08-04						
PRIOR APPLICATION NUMBER: AU PQ7761/00						
PRIOR FILING DATE: 2000-05-26						
NUMBER OF SEQ ID NOS: 1507						
SOFTWARE: PatentIn version 3.2						
SEQ ID NO 406						
LENGTH: 3011						
TYPE: PRT						
ORGANISM: Artificial						
FEATURE:						
OTHER INFORMATION: HepC 1a consensus polyprotein						
US-10-296-734-406						
Query Match	Best Local Similarity	100.0%;	Score 20;	DB 6;	Length 3011;	
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	WHYPPKPGGIVPAKSVCGPV 20				
DB	487	WHYPPKPGGIVPAKSVCGPV 506				
RESULT 7						
PCT-US03-19834-2						
Sequence 2, Application PC/TUS0319834						
GENERAL INFORMATION:						
APPLICANT: STAPLETON, JACK T.						
APPLICANT: WUENSCHMANN, SABINA						
TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND METHODS FOR TREATING HYPERLIPIDEMIA						
FILE REFERENCE: IOWA:045WO						
CURRENT APPLICATION NUMBER: PCT/US03/19834						
PRIOR FILING DATE: 2003-06-24						
PRIOR APPLICATION NUMBER: 10/445,724						
PRIOR FILING DATE: 2003-05-27						
PRIOR APPLICATION NUMBER: 60/392,158						
PRIOR FILING DATE: 2002-06-28						
NUMBER OF SEQ ID NOS: 3						
SOFTWARE: PatentIn Ver. 2.1						
SEQ ID NO 2						
LENGTH: 3011						
TYPE: PRT						
ORGANISM: Hepatitis C virus						
PCT-US03-19834-2						
Query Match	Best Local Similarity	100.0%;	Score 20;	DB 1;	Length 3011;	
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	WHYPPKPGGIVPAKSVCGPV 20				
DB	487	WHYPPKPGGIVPAKSVCGPV 506				

Query Match	Best Local Similarity	Score	DB	Length	Indels	Mismatches	Gaps
<p>RESULT 7</p> <p>PC-TUS03-19834-2</p> <p>Sequence 2, Application PC/TUS0319834</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: STAPLETON, JACK T.</p> <p>APPLICANT: WUENSCHMANN, SABINA</p> <p>TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND METHODS FOR TREATING HYPERLIPIDEMIA</p> <p>FILE REFERENCE: IOWA:045WO</p> <p>CURRENT APPLICATION NUMBER: PCT/US03/19834</p> <p>CURRENT FILING DATE: 2003-06-24</p> <p>PRIOR APPLICATION NUMBER: 10/445,724</p> <p>PRIOR FILING DATE: 2003-05-27</p> <p>PRIOR APPLICATION NUMBER: 60/392,158</p> <p>PRIOR FILING DATE: 2002-06-28</p> <p>NUMBER OF SEQ ID NOS: 3</p> <p>SOFTWARE: PatentIn Ver. 2.1</p> <p>SEQ ID NO 2</p> <p>LENGTH: 3011</p> <p>TYPE: PRT</p> <p>ORGANISM: Hepatitis C virus</p> <p>PC-TUS03-19834-2</p>							
Query Match	100.0%	Score 20;	DB 1;	Length 3011;	Indels 0;	Mismatches 0;	Gaps 0;
Best Local Similarity	100.0%	Pred. No. 6.7e-12;					
Matches	20;	Conservative 0;					
QY	1	WHYPPKPGGIVPAKSVCGPV 20					
DB	487	WHYPPKPGGIVPAKSVCGPV 506					
<p>RESULT 8</p> <p>US-10-296-734-406</p> <p>Sequence 406, Application US/10296734</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Thompson, Scott A</p> <p>APPLICANT: Ramshaw, Ian A</p> <p>TITLE OF INVENTION: Synthetic molecules and uses therefor</p> <p>FILE REFERENCE: Savine</p> <p>CURRENT APPLICATION NUMBER: US/10/296,734</p> <p>CURRENT FILING DATE: 2003-08-04</p> <p>PRIOR APPLICATION NUMBER: AU PQ7761/00</p> <p>PRIOR FILING DATE: 2000-05-26</p> <p>NUMBER OF SEQ ID NOS: 1507</p> <p>SOFTWARE: PatentIn version 3.2</p> <p>SEQ ID NO 406</p> <p>LENGTH: 3011</p> <p>TYPE: PRT</p> <p>ORGANISM: Artificial</p> <p>FEATURE:</p> <p>OTHER INFORMATION: HepC 1a consensus polypeptide</p> <p>US-10-296-734-406</p>							
Query Match	100.0%	Score 20;	DB 6;	Length 3011;	Indels 0;	Mismatches 0;	Gaps 0;
Best Local Similarity	100.0%	Pred. No. 6.7e-12;					
Matches	20;	Conservative 0;					
QY	1	WHYPPKPGGIVPAKSVCGPV 20					
DB	487	WHYPPKPGGIVPAKSVCGPV 506					
<p>RESULT 9</p> <p>US-10-296-734-810</p> <p>Sequence 810, Application US/10296734</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Thompson, Scott A</p> <p>APPLICANT: Ramshaw, Ian A</p> <p>TITLE OF INVENTION: Synthetic molecules and uses therefor</p> <p>FILE REFERENCE: Savine</p> <p>CURRENT APPLICATION NUMBER: US/10/296,734</p>							

Query Match	Best Local Similarity	Score	DB	Length	Indels	Mismatches	Gaps
<p>RESULT 7</p> <p>PC-TUS03-19834-2</p> <p>Sequence 2, Application PC/TUS0319834</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: STAPLETON, JACK T.</p> <p>APPLICANT: WUENSCHMANN, SABINA</p> <p>TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND METHODS FOR TREATING HYPERLIPIDEMIA</p> <p>FILE REFERENCE: IOWA:045WO</p> <p>CURRENT APPLICATION NUMBER: PCT/US03/19834</p> <p>CURRENT FILING DATE: 2003-06-24</p> <p>PRIOR APPLICATION NUMBER: 10/445,724</p> <p>PRIOR FILING DATE: 2003-05-27</p> <p>PRIOR APPLICATION NUMBER: 60/392,158</p> <p>PRIOR FILING DATE: 2002-06-28</p> <p>NUMBER OF SEQ ID NOS: 3</p> <p>SOFTWARE: PatentIn Ver. 2.1</p> <p>SEQ ID NO 2</p> <p>LENGTH: 3011</p> <p>TYPE: PRT</p> <p>ORGANISM: Hepatitis C virus</p> <p>PC-TUS03-19834-2</p>							
Query Match	100.0%	Score 20;	DB 1;	Length 3011;	Indels 0;	Mismatches 0;	Gaps 0;
Best Local Similarity	100.0%	Pred. No. 6.7e-12;					
Matches	20;	Conservative 0;					
QY	1	WHYPPKPGGIVPAKSVCGPV 20					
DB	487	WHYPPKPGGIVPAKSVCGPV 506					
<p>RESULT 8</p> <p>US-10-296-734-406</p> <p>Sequence 406, Application US/10296734</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Thompson, Scott A</p> <p>APPLICANT: Ramshaw, Ian A</p> <p>TITLE OF INVENTION: Synthetic molecules and uses therefor</p> <p>FILE REFERENCE: Savine</p> <p>CURRENT APPLICATION NUMBER: US/10/296,734</p> <p>CURRENT FILING DATE: 2003-08-04</p> <p>PRIOR APPLICATION NUMBER: AU PQ7761/00</p> <p>PRIOR FILING DATE: 2000-05-26</p> <p>NUMBER OF SEQ ID NOS: 1507</p> <p>SOFTWARE: PatentIn version 3.2</p> <p>SEQ ID NO 406</p> <p>LENGTH: 3011</p> <p>TYPE: PRT</p> <p>ORGANISM: Artificial</p> <p>FEATURE:</p> <p>OTHER INFORMATION: HepC 1a consensus polypeptide</p> <p>US-10-296-734-406</p>							
Query Match	100.0%	Score 20;	DB 6;	Length 3011;	Indels 0;	Mismatches 0;	Gaps 0;
Best Local Similarity	100.0%	Pred. No. 6.7e-12;					
Matches	20;	Conservative 0;					
QY	1	WHYPPKPGGIVPAKSVCGPV 20					
DB	487	WHYPPKPGGIVPAKSVCGPV 506					
<p>RESULT 9</p> <p>US-10-296-734-810</p> <p>Sequence 810, Application US/10296734</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Thompson, Scott A</p> <p>APPLICANT: Ramshaw, Ian A</p> <p>TITLE OF INVENTION: Synthetic molecules and uses therefor</p> <p>FILE REFERENCE: Savine</p> <p>CURRENT APPLICATION NUMBER: US/10/296,734</p>							

	Best Local Similarity	100.0%;	Pred. No. 1;	Mismatches	7; Conservative	0; Indels	0; Gaps	0;
QY	1 WHYPKPKP	7						
DB	24 WHYPKPKP	30						
 RESULT 14								
US-10-685-435-21								
; Sequence 21, Application US/10685435								
; GENERAL INFORMATION:								
; APPLICANT: Imogenetics N.V.								
; TITLE OF INVENTION: Multi-mer peptides derived from Hepatitis C Virus								
; TITLE OF INVENTION: envelope proteins for diagnostic use and vaccination								
; TITLE OF INVENTION: purposes.								
; FILE REFERENCE: PCT98.75.HCV30								
; CURRENT APPLICATION NUMBER: US/10/685.435								
; CURRENT FILING DATE: 2003-10-16								
; PRIOR APPLICATION NUMBER: US/09/566,266B								
; PRIOR FILING DATE: 2000-05-05								
; PRIOR APPLICATION NUMBER: 97870179.5								
; PRIOR FILING DATE: 1997-11-06								
; NUMBER OF SEQ ID NOS: 41								
; SOFTWARE: patenting ver. 2.1								
; SEQ ID NO 21								
; LENGTH: 34								
; TYPE: PRT								
; ORGANISM: Hepatitis C virus								
US-10-685-435-21								
 Query Match								
Best Local Similarity								
Mismatches								
Score 7; DB 6; Length 34;								
Pred. No. 1.1;								
Conservative 0; Mismatches 0; Indels 0; Gaps 0;								
QY	7 PCGIVPA	13						
DB	14 PCGIVPA	20						
 RESULT 15								
US-10-664-391-9								
; Sequence 9, Application US/10664391								
; GENERAL INFORMATION:								
; APPLICANT: Donnelly, John J.								
; APPLICANT: Liu, Margaret A.								
; APPLICANT: Shiver, John W.								
; APPLICANT: Fu, Tong-Ming								
; TITLE OF INVENTION: SYNTHETIC HEPATITIS C GENES								
; FILE REFERENCE: 19732YPCA								
; CURRENT APPLICATION NUMBER: US/10/664,391								
; CURRENT FILING DATE: 2003-09-17								
; PRIOR APPLICATION NUMBER: PCT/US97/09884								
; PRIOR FILING DATE: 1997-06-06								
; PRIOR APPLICATION NUMBER: 60/033,534								
; PRIOR FILING DATE: 1996-12-20								
; PRIOR APPLICATION NUMBER: 60/020,494								
; PRIOR FILING DATE: 1996-06-11								
; NUMBER OF SEQ ID NOS: 25								
; SOFTWARE: FastSeq for Windows Version 4.0								
; SEQ ID NO 9								
; LENGTH: 347								
; TYPE: PRT								
; ORGANISM: Hepatitis C Virus								
US-10-664-391-9								
 Query Match								
Best Local Similarity								
Mismatches								
Score 7; DB 6; Length 347;								
Pred. No. 5.3;								
Conservative 0; Mismatches 0; Indels 0; Gaps 0;								
QY	7 PCGIVPA	13						
DB	111 PCGIVPA	117						

Search completed: November 21, 2003, 22:12:55
Job time : 8.55 secs

0

0

0

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:46:27 ; Search time 9.25 Seconds
(without alignments)
207.932 Million cell updates/sec

Title: US-09-973-025-79

Perfect score: 20

Sequence: 1 WHYPKPCGIVPAKSVCGPV 20

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	3011	1 GNWVC3	genome polyprotein
2	20	100.0	3011	1 S40770	genome polyprotein
3	14	70.0	138	2 S24081	envelope protein -
4	14	70.0	3011	1 GNWVCH	genome polyprotein
5	13	65.0	640	2 JQ1584	genome polyprotein
6	13	65.0	716	2 JQ1366	polyprotein - hepa
7	11	55.0	513	2 A44150	structural protein
8	11	55.0	874	2 JQ0883	genome polyprotein
9	10	50.0	138	2 S24080	envelope protein -
10	9	45.0	138	2 S24075	envelope protein -
11	9	45.0	138	2 S24074	envelope protein -
12	9	45.0	138	2 S24092	envelope protein -
13	9	45.0	787	2 PN0677	hypothetical prote
14	7	35.0	138	2 PC1205	envelope protein -
15	7	35.0	138	2 PC1203	envelope protein -
16	7	35.0	138	2 PC1206	envelope protein -
17	7	35.0	138	2 PC1207	envelope protein -
18	7	35.0	138	2 PC1208	envelope protein -
19	7	35.0	138	2 PC1209	envelope protein -
20	7	35.0	138	2 PC1210	envelope protein -
21	7	35.0	138	2 PC1211	envelope protein -
22	7	35.0	138	2 PC1212	envelope protein -
23	7	35.0	138	2 PC1197	genome polyprotein
24	7	35.0	138	2 PC1199	genome polyprotein
25	7	35.0	138	2 PC1200	genome polyprotein
26	7	35.0	138	2 PC1201	envelope protein -
27	7	35.0	138	2 PC1204	envelope protein -
28	7	35.0	138	2 S24070	envelope protein -
29	7	35.0	138	2 S24076	envelope protein -

30 7 35.0 138 2 S24095 envelope protein -
31 7 35.0 138 2 S24102 envelope protein -
32 7 35.0 138 2 S24099 envelope protein -
33 7 35.0 138 2 S24085 envelope protein -
34 7 35.0 138 2 S24083 envelope protein -
35 7 35.0 138 2 S24079 envelope protein -
36 7 35.0 138 2 S24105 envelope protein -
37 7 35.0 138 2 S24077 envelope protein -
38 7 35.0 138 2 S24106 envelope protein -
39 7 35.0 138 2 S24094 envelope protein -
40 7 35.0 138 2 S24071 envelope protein -
41 7 35.0 138 2 S24104 envelope protein -
42 7 35.0 138 2 S24082 envelope protein -
43 7 35.0 138 2 S24101 envelope protein -
44 7 35.0 138 2 S24103 envelope protein -
45 7 35.0 139 2 S24072 envelope protein -

ALIGNMENTS

RESULT 1

GNWVC3

genome polyprotein - hepatitis C virus (strain HCV-1)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 19-Jan-2001
C:Accession: A39166; PQ0403; PQ0404
R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Co Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A:Title: Genetic organization and diversity of the hepatitis C virus.
A:Reference number: A39166; MUID:91172826; PMID:1848704
A:Accession: A39166
A:Molecule type: mRNA
A:Residues: 1-3011 <CHO>
A:Cross-References: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.I J. Gen. Virol. 73, 1131-1141, 1992
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to A:Reference number: PQ0393; MUID:92268871; PMID:1316939
A:Accession: PQ0403
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CHA>
A:Cross-References: DBJ:D10128
A:Experimental source: isolates E-b16
A:Accession: PQ0404
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CH2>
A:Experimental source: isolates E-b17
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein M; capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 100.0%; Score 20; DB 1; Length 3011;

Best Local Similarity 100.0%; Pred. No. 2.2e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPKPCGIVPAKSVCGPV 20

|||||

Db 487 WHYPPKCGIVPAKSVCGPV 506

RESULT 2

S40770

genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: S40770; PC1285
R:Okamoto, H.
submitted to the EMBL Data Library, March 1992
A:Reference number: S40770
A:Accession: S40770
A:Molecule type: genomic RNA
A:Residues: 1-3011 <OKA>
A:Cross-references: EMBL:D10749; NID:G221586; PIDN:BAA01582.1; PID:G221587
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jun. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; MUID:91013116; PMID:2170712
A:Accession: PC1285
A:Molecule type: genomic RNA
A:Residues: 1-513 <OK2>
A:Cross-references: GB:D00831; NID:G221511; PIDN:BAA00705.1; PID:G221512
A:Experimental source: isolate HC-J1
A:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepacivirin #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.2e-13; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

QY 1 WHYPPKCGIVPAKSVCGPV 20
|||||
Db 487 WHYPPKCGIVPAKSVCGPV 506

RESULT 3

S24081

envelope protein - hepatitis C virus (fragment)
N:Contains: envelope protein gp35; envelope protein gp70
C:Species: hepatitis C virus
C:Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C:Accession: S24081
R:Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; Muraio, K.; Ohkoshi, Virus Res. 22, 107-123, 1992
A:Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A:Reference number: S24068; MUID:92230394; PMID:1314471
A:Accession: S24081
A:Molecule type: genomic RNA
A:Residues: 1-138 <KAT>
A:Cross-references: EMBL:X60569
A:Experimental source: isolate R243
A:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; glycoprotein; polyprotein
F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted

F:49,55,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 70.0%; Score 14; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 5.1e-08; Mismatches 0; Indels 0; Gaps 0;
Matches 14; Conservative 0;

QY 7 PCGIVPAKSVCGPV 20
|||||
Db 125 PCGIVPAKSVCGPV 138

RESULT 4

GNVCH

genome polyprotein - hepatitis C virus (strain H)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A36814; A41546
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: C
A:Reference number: A36814
A:Accession: A36814
A:Molecule type: genomic RNA
A:Residues: 1-3011 <INC>
A:Cross-references: GB:M67463; NID:G329737; PIDN:AAA45534.1; PID:G329738
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: comparison
A:Reference number: A41546; MUID:92052256; PMID:1658800
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepacivirin #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match 70.0%; Score 14; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 5.5e-07; Mismatches 0; Indels 0; Gaps 0;
Matches 14; Conservative 0;

QY 7 PCGIVPAKSVCGPV 20
|||||
Db 493 PCGIVPAKSVCGPV 506

RESULT 5

JQ1584

genome polyprotein - hepatitis C virus (strain U.K.) (fragment)
N:Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural pr
C:Species: hepatitis C virus
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
C:Accession: JQ1584
R:Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.
J. Gen. Virol. 73, 1521-1525, 1992
A:Title: Cloning and sequencing of the structural region and expression of putative core
A:Reference number: JQ1584; MUID:92300349; PMID:1318944
A:Accession: JQ1584
A:Molecule type: genomic RNA
A:Residues: 1-640 <KUM>

A;Cross-references: GB:X84079; NID:G643119; PIDN:CAA58888.1; PID:G643120
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; polyprotein
 F:1-191/Product: core protein C #status predicted <CPC>
 F:192-389/Product: envelope protein E1 #status predicted <SE1>
 F:390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predicted <NS1>
 F:196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn) (covar)

Query Match 65.0%; Score 13; DB 2; Length 640;
 Best Local Similarity 100.0%; Pred. No. 2e-06;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPCGIVPA 13
 |||||
 DB 487 WHYPPKPCGIVPA 499

RESULT 6
 JQ1366
 polyprotein - hepatitis C virus (French isolate) (fragments)
 C;Species: hepatitis C virus
 C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
 C;Accession: JQ1366
 R;Kremsdorff, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.
 J. Gen. Virol. 72, 2557-2561, 1991
 A;Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication
 A;Reference number: JQ1366; MUID:92013977; PMID:1655961
 A;Accession: JQ1366
 A;Molecule type: genomic RNA
 A;Residues: 1-716 <KRE>
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: glycoprotein; polyprotein
 F:84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #stat

Query Match 65.0%; Score 13; DB 2; Length 716;
 Best Local Similarity 100.0%; Pred. No. 2.1e-06;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPCGIVPA 13
 |||||
 DB 154 WHYPPKPCGIVPA 166

RESULT 7
 A44150
 structural protein - hepatitis C virus
 C;Species: hepatitis C virus
 C;Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 17-Nov-2000
 C;Accession: A44150
 R;Ching, W.M.; Wychowski, C.; Beach, M.J.; Wang, H.; Davies, C.L.; Carl, M.; Bradley, D.
 Proc. Natl. Acad. Sci. U.S.A. 89, 3190-3194, 1992
 A;Title: Interaction of immune sera with synthetic peptides corresponding to the structu
 A;Reference number: A44150; MUID:92228749; PMID:1373489
 A;Accession: A44150
 A;Status: preliminary; not compared with conceptual translation
 A;Molecule type: genomic RNA
 A;Residues: 1-513 <CHI>
 C;Superfamily: hepatitis C virus genome polyprotein

Query Match 55.0%; Score 11; DB 2; Length 513;
 Best Local Similarity 100.0%; Pred. No. 0.00022;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 IVPKSVCGPV 20
 |||||
 DB 496 IVPKSVCGPV 506

RESULT 8
 JQ0883
 genome polyprotein - hepatitis C virus (strain J7) (fragments)
 N;Contains: NS5 protein
 C;Species: hepatitis C virus

C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 23-Mar-2001
 C;Accession: JQ0883
 R;Okamoto, H.
 submitted to JPIID, January 1991
 A;Description: The 5'-terminal and 3'-terminal sequences of the genomic RNA of hepatitis
 A;Reference number: JQ0879
 A;Accession: JQ0883
 A;Molecule type: genomic RNA
 A;Residues: 1-874 <OKA>
 A;Experimental source: strain J7
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: polyprotein
 F:510-874/Product: NS5 protein (fragment) #status predicted <NS5>

Query Match 55.0%; Score 11; DB 2; Length 874;
 Best Local Similarity 100.0%; Pred. No. 0.00034;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPCGIV 11
 |||||
 DB 489 WHYPPKPCGIV 499

RESULT 9
 S24080
 envelope protein - hepatitis C virus (fragment)
 N;Contains: envelope protein gp35; envelope protein gp70
 C;Species: hepatitis C virus
 C;Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
 C;Accession: S24080
 R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; MuraIso, K.; Ohkoshi,
 Virus Res. 22, 107-123, 1992
 A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C vir
 A;Reference number: S24068; MUID:92230394; PMID:1314471
 A;Accession: S24080
 A;Molecule type: genomic RNA
 A;Residues: 1-138 <KAT>
 A;Cross-references: EMBL:X60568
 A;Experimental source: isolate RE42
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: envelope protein; glycoprotein; polyprotein
 F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
 F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
 F:16-42/Region: hypervariable 1 #status predicted
 F:106-112/Region: hypervariable 2 #status predicted
 F:49,55,62,80/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 50.0%; Score 10; DB 2; Length 138;
 Best Local Similarity 100.0%; Pred. No. 0.00094;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPCGI 10
 |||||
 DB 119 WHYPPKPCGI 128

RESULT 10
 S24075
 envelope protein - hepatitis C virus (fragment)
 N;Contains: envelope protein gp35; envelope protein gp70
 C;Species: hepatitis C virus
 C;Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
 C;Accession: S24075
 R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; MuraIso, K.; Ohkoshi,
 Virus Res. 22, 107-123, 1992
 A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C vir
 A;Reference number: S24068; MUID:92230394; PMID:1314471
 A;Accession: S24075
 A;Molecule type: genomic RNA
 A;Residues: 1-138 <KAT>
 A;Cross-references: EMBL:X60564
 A;Experimental source: isolate RE37B
 C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: envelope protein; glycoprotein; polyprotein
F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted
F:49,55,62,80,108/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 45.0%; Score 9; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPPKPCG 9
Db 119 WHYPPKPCG 127

RESULT 11
S24074
envelope protein - hepatitis C virus (fragment)
N;Contains: envelope protein gp35; envelope protein gp70
C;Species: hepatitis C virus
C;Accession: S24074
C;Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; MuraIso, K.; Ohkoshi,
Virus Res. 22, 107-123, 1992
A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A;Reference number: S24068; MUID:92230394; PMID:1314471
A;Accession: S24074
A;Molecule type: genomic RNA
A;Residues: 1-138 <KAT>
A;Cross-references: EMBL:X60563
A;Experimental source: isolate RE37A
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: envelope protein; glycoprotein; polyprotein
F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted
F:49,55,62,80,108/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 45.0%; Score 9; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WHYPPKPCG 9
Db 119 WHYPPKPCG 127

RESULT 12
S24092
envelope protein - hepatitis C virus (fragment)
N;Contains: envelope protein gp35; envelope protein gp70
C;Species: hepatitis C virus
C;Date: 22-Nov-1993 #sequence_revision 13-Feb-1998 #text_change 17-Nov-2000
C;Accession: S24092
R;Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M.; Nakazawa, T.; MuraIso, K.; Ohkoshi,
Virus Res. 22, 107-123, 1992
A;Title: Marked sequence diversity in the putative envelope proteins of hepatitis C virus
A;Reference number: S24068; MUID:92230394; PMID:1314471
A;Accession: S24092
A;Molecule type: genomic RNA
A;Residues: 1-138 <KAT>
A;Cross-references: EMBL:X60578
A;Experimental source: isolate RE56
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: envelope protein; glycoprotein; polyprotein
F:1-15/Product: envelope protein gp35 (fragment) #status predicted <MAT1>
F:16-138/Product: envelope protein gp70 (fragment) #status predicted <MAT2>
F:16-42/Region: hypervariable 1 #status predicted
F:106-112/Region: hypervariable 2 #status predicted
F:49,55,62,80,110/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 45.0%; Score 9; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 PKPCGIVPA 13
Db 123 PKPCGIVPA 131

RESULT 13
PN0677
hypothetical protein 787 - hepatitis C virus (fragment)
C;Species: hepatitis C virus
C;Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000
C;Accession: PN0677
R;Cho, S.H.; Yoon, J.I.; Chang, J.E.; Ahn, B.M.; Lee, C.H.; Lee, Y.I.
Biochem. Biophys. Res. Commun. 196, 780-788, 1993
A;Title: Genomic typing of hepatitis C viruses from Korean patients: Implications of gen
A;Reference number: PN0677; MUID:94059104; PMID:8240354
A;Accession: PN0677
A;Molecule type: mRNA
A;Residues: 1-787 <CHO>
A;Cross-references: GB:L20498; NID:gl381031; PIDN:AAB02608.1; PID:gl381032
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: glycoprotein; nonstructural protein
F:196,209,234,250,305,325,421,427,452,536,544,560,580,627,649/Binding site: carbohydrate

Query Match 45.0%; Score 9; DB 2; Length 787;
Best Local Similarity 100.0%; Pred. No. 0.042;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 PKPCGIVPA 13
Db 495 PKPCGIVPA 503

RESULT 14
PC1205
envelope protein - hepatitis C virus (strain RS3-2) (fragment)
C;Species: hepatitis C virus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C;Accession: PC1205
R;Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Sekiya, H.; Hijikata, M.; Shimotoch
Biochem. Biophys. Res. Commun. 189, 119-127, 1992
A;Title: Characterization of hypervariable regions in the putative envelope protein of h
A;Reference number: PC1182; MUID:93080545; PMID:1333186
A;Accession: PC1205
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-138 <KAT>
A;Cross-references: DDBJ:D12942; DDBJ:D12972
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: envelope protein

Query Match 35.0%; Score 7; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 PCGIVPA 13
Db 125 PCGIVPA 131

RESULT 15
PC1203
envelope protein - hepatitis C virus (strain RS2-4) (fragment)
C;Species: hepatitis C virus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C;Accession: PC1203
R;Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Sekiya, H.; Hijikata, M.; Shimotoch
Biochem. Biophys. Res. Commun. 189, 119-127, 1992
A;Title: Characterization of hypervariable regions in the putative envelope protein of h
A;Reference number: PC1182; MUID:93080545; PMID:1333186

A:Accession: PC1203
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-138 <KAT>
A:Cross-references: DDBJ:DL2942; DDBJ:DL2972
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein

Query Match 35.0%; Score 7; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 PCGIVPA 13
DB 125 PCGIVPA 131

Search completed: November 21, 2003, 21:11:31
Job time : 9.25 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:52 ; Search time 4.85 seconds
(without alignments)
193.925 Million cell updates/sec

Title: US-09-973-025-79

Perfect score: 20

Sequence: 1 WHYPPKPGIVPAKSVCGPV 20

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	3011	1 POLG_HCV1	P26664 h genome po
2	14	70.0	3011	1 POLG_HCVH	P27958 h genome po
3	13	65.0	737	1 POLG_HCVJ5	P27960 hepatitis c
4	11	55.0	737	1 POLG_HCVJ7	P27961 hepatitis c
5	7	35.0	3010	1 POLG_HCVJA	P26662 h genome po
6	7	35.0	3010	1 POLG_HCVTW	P2846 h genome po
7	7	35.0	3033	1 POLG_HCVJ8	P26661 h genome po
8	6	30.0	62	1 PLR_RANPI	Q90wp7 rana pipien
9	6	30.0	183	1 YRDC_HAEIN	P44807 haemophilus
10	6	30.0	275	1 IL2A_FELCA	P41690 felis silve
11	6	30.0	285	1 KC2C_NEUCR	Q8tg11 neurospora
12	6	30.0	454	1 COBB_THEVO	Q97bp4 thermoplas
13	6	30.0	454	1 SDHL_ECOLI	P16095 escherichia
14	6	30.0	1274	1 ENAM_MOUSE	O55196 mus musculu
15	6	30.0	3010	1 POLG_HCVJT	Q00269 h genome po
16	6	30.0	3033	1 POLG_HCVJ6	P26660 h genome po
17	5	25.0	36	1 PYV_AMICA	P29205 amia calva
18	5	25.0	36	1 PYV_LEPSP	P09473 lepisosteus
19	5	25.0	36	1 PYV_ONCKI	P09474 oncorhynch
20	5	25.0	36	1 PYV_ORENI	P81028 oreochromis
21	5	25.0	36	1 PYV_RAJRH	P29206 raja rhina
22	5	25.0	36	1 PYV_RANRI	P29204 rana ridibu
23	5	25.0	36	1 SPVY_PHYBI	P80952 phyllomedus
24	5	25.0	37	1 PYV_CHICK	P29203 gallus gall
25	5	25.0	69	1 PYV_LOPAM	P09475 lophius ame
26	5	25.0	91	1 YAS7_STAAM	Q99V05 staphylococ
27	5	25.0	96	1 CH10_CANTP	Q8ktr9 candidatus
28	5	25.0	97	1 PYV_BRARE	Q918p2 brachydanio
29	5	25.0	97	1 PYV_DICLA	Q9pt98 dicentrarch
30	5	25.0	122	1 CASK_MAZAM	Q95191 mazama amer
31	5	25.0	122	1 CASK_ODOHE	Q95225 odocoileus
32	5	25.0	122	1 CASK_ODOVI	Q95228 odocoileus
33	5	25.0	122	1 CASK_RANTA	Q95239 rangifer ta

RESULT 1

ID	POLG_HCV1	STANDARD;	PRT;	3011 AA.
AC	P26664;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Genome Polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)]			
OS	Hepatitis C virus (isolate 1) (HCV)			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus			
OX	NCBI_TaxID=11104;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91172826; PubMed=1848704;			
RA	Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;			
RA	"Genetic organization and diversity of the hepatitis C virus.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).			
CC	!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.			
CC	!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA) (N).			
CC	!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.			
CC	!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; M62321; AAA45676.1; -			
CC	PIR; A39166; GNMVCS3.			
CC	PDB; 1A1V; 16-FEB-99.			
CC	PDB; 1HEI; 25-NOV-98.			
CC	MEROPS; S29.001; -			
CC	MEROPS; U39.001; -			
CC	InterPro; IPR001410; DEAD.			
CC	InterPro; IPR002522; HCV_capsid.			

ALIGNMENTS

34	5	25.0	123	1	CASK_CAPCA	Q95146 capreolus c
35	5	25.0	124	1	CASK_OVIDA	Q95224 ovis dalli
36	5	25.0	136	1	CASK_BISBO	P42155 bison bonas
37	5	25.0	143	1	MCS_MOUSE	P15265 mus musculu
38	5	25.0	143	1	MRAZ_BACSU	P53343 bacillus su
39	5	25.0	143	1	MRAZ_LISIN	Q929X5 listeria in
40	5	25.0	143	1	MRAZ_LISMO	Q8Y516 listeria mo
41	5	25.0	145	1	MCS_RAT	Q64298 rattus norv
42	5	25.0	153	1	CASK_GIRCA	Q28417 giraffa cam
43	5	25.0	162	1	RSBW_BACHD	Q9kff1 bacillus ha
44	5	25.0	165	1	COAD_BRAJA	Q89155 bradyrhizob
45	5	25.0	172	1	YKKB_BACSU	P49855 bacillus su

DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RDRP.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR007095; RNA pol DS ps.
 DR InterPro; IPR007094; RNA pol psivir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV NS1; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; Helicase C; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3011
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1093
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 136 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 476 476
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2364 2364
 FT CARBOHYD 2789 2789
 FT SEQUENCE 3011 AA; 32171 MW; 65F8C9447FCE5A99 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 5.9e-14;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WHYPPKCGIVPAKSVCGPV 20

Db 487 WHYPPKCGIVPAKSVCGPV 506
 RESULT 2 0
 POLG_HCVH STANDARD; PRT; 3011 AA.
 ID POLG_HCVH
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=1658800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding";
 RL Structure 6:89-100(1998).
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position. Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS; E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MENA.
 CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.

CC EMBL; M67463; AAA45534.1; .
DR PIR; A36814; GNVVCH.
DR PDB; IHEI; 25-NOV-98.
DR PDB; IAI; 16-FEB-99.
DR PDB; IAI; 17-JUN-98.
DR MEROPS; S29.001; .
DR MEROPS; U39.001; .
DR TRANSFAC; T04155; .
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR002518; HCV NS3.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA pol DS PS.
DR InterPro; IPR007094; RNA pol PSvir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; helicase C; 1.
DR Pfam; PF00998; Viral RdRp; 1.
DR ProDom; PDI86062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
FT CHAIN 192 383 CAPSID PROTEIN C.
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
FT CHAIN 810 1026 PROTEIN P7.
FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5A.
FT CHAIN 3011 369 NONSTRUCTURAL PROTEIN NS5B.
FT TRANSMEM 347 369 POTENTIAL.
FT ACT SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 1107 1107 ACT SITE (BY SIMILARITY).
FT ACT SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DESCH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 224 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT STRAND 1224 1226
FT TURN 1232 1233
FT TURN 1236 1238
FT TURN 1239 1246
FT TURN 1247 1248
FT STRAND 1251 1255
FT TURN 1258 1271
FT TURN 1272 1272
FT STRAND 1277 1280
FT TURN 1281 1282
FT STRAND 1283 1285
FT STRAND 1291 1295
FT HELIX 1296 1301
FT TURN 1302 1303
FT STRAND 1312 1316
FT TURN 1317 1319
FT TURN 1323 1335
FT TURN 1336 1340
FT STRAND 1343 1347
FT TURN 1352 1353
FT TURN 1360 1361
FT STRAND 1362 1366
FT STRAND 1368 1368
FT STRAND 1373 1375
FT TURN 1376 1377
FT STRAND 1378 1380
FT HELIX 1382 1385
FT STRAND 1389 1393
FT TURN 1397 1409
FT TURN 1410 1411
FT STRAND 1414 1417
FT TURN 1419 1420
FT STRAND 1432 1436
FT TURN 1438 1439
FT STRAND 1450 1453
FT STRAND 1456 1463
FT STRAND 1471 1478
FT STRAND 1480 1480
FT HELIX 1481 1488
FT TURN 1489 1490
FT STRAND 1497 1501
FT STRAND 1507 1507
FT STRAND 1511 1511
FT HELIX 1514 1527
FT TURN 1532 1544
FT STRAND 1550 1550
FT HELIX 1555 1564
FT TURN 1570 1578
FT TURN 1579 1580
FT TURN 1584 1597
FT TURN 1598 1598
FT TURN 1606 1611
FT TURN 1614 1618
FT STRAND 1622 1623
FT STRAND 1627 1627
FT STRAND 1635 1636
FT HELIX 1640 1652
SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;

Query Match 70.0%; Score 14; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 PCGIVPAKSVCGPV 20
Db 493 PCGIVPAKSVCGPV 506

RESULT 3
POLG_HCVJ5
ID POLG_HCVJ5 STANDARD; PRT; 737 AA.
AC P27960; AC P27960;
DT 01-AUG-1992 (Rel. 23, Created)

FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327017 MW; AA93794F46DB185 CRC64;

Query Match 35.0%; Score 7; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 PGIIVPA 13
DB 493 PGIIVPA 499

RESULT 6
POLG_HCVTW
ID POLG_HCVTW STANDARD; PRT; 3010 AA.
AC P29846;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/Helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Taiwan) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31645;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230206; PubMed=1314449;
RA Chen P.J., Lin M.H., Tai K.P., Liu P.C., Lin C.J., Chen D.S.;
RT "The Taiwanese hepatitis C virus genome: sequence determination and
RT mapping the 5' terminus of viral genomic and antigenomic RNA";
RL Virology 188:102-113(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M84754; -; NOT ANNOTATED_CDS.
DR PIR; A40244; GNWVTV.
DR PDB; 1N64; 25-PEB-03.
DR PDB; 1NS3; 08-APR-98.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.

DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RDRP.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RDRP; 1.
DR Pfam; PF0186062; HCV_NS1; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 CORE PROTEIN (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT ACT_SITE 1083 1083 POTENTIAL.
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 35.0%; Score 7; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 PGIIVPA 13

DB 493 PGIIVPA 499

RESULT 7

POLG_HCVJ8

ID AC P26661; POLG_HCVJ8 STANDARD; PRT; 3033 AA.
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J8) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11115;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
 RA Fukuda S., Tenda F., Mishihiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GP24.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D10988; BAA01761.1; -;
 DR PIR; A40250; GNMVJ8.
 DR HSSP; P27958; 1HEI.
 DR MEROPS; S29.001; -;
 DR MEROPS; U39.001; -;
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4A.
 DR InterPro; IPR001490; HCV NS4B.
 DR InterPro; IPR002868; HCV NS5A.
 DR InterPro; IPR002166; HCV RdRP.
 DR InterPro; IPR007095; RNA pol DS PS.
 DR InterPro; IPR007094; RNA pol PSvir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.

DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDc; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2359 2359
 FT CARBOHYD 2811 2811
 SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E381FD1A CRC64;
 Query Match 35.0%; Score 7; DB 1; Length 3033;
 Best Local Similarity 100.0%; Pred. No. 6.5;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 PCGIVPA 13
 DB 495 PCGIVPA 501
 RESULT 8
 PLR_RANPI
 ID PLR_RANPI STANDARD; PRT; 62 AA.
 AC Q90WP7;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Peptide leucine arginine precursor (pLR).
 OS Rana pipiens (Northern leopard frog).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Rana.
 OX NCBI_TaxID=8404;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin;
 RA Farragher S., Bjorson A.J., McClean S., Orr D.F., Shaw C.;
 RT "Cloning of cDNAs encoding defensive skin secretion peptides from the
 RT Northern leopard frog (Rana pipiens).";

FT CHAIN 22 275 INTERLEUKIN-2 RECEPTOR ALPHA CHAIN.
 FT DOMAIN 22 243 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 244 262 POTENTIAL.
 FT DOMAIN 263 275 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 23 78 SUSHI 1.
 FT DOMAIN 122 185 SUSHI 2.
 FT DISULFID 24 67 BY SIMILARITY.
 FT DISULFID 751 80 BY SIMILARITY.
 FT DISULFID 125 168 BY SIMILARITY.
 FT DISULFID 152 184 BY SIMILARITY.
 FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 70 70 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 275 AA; 30817 MW; C97B8D8D0632326 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 275;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 PAKSVC 17
 DB 163 PAKSVC 168

RESULT 11

ID KC2C NEUCR STANDARD; PRT; 285 AA.
 AC Q8TG11;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Casein kinase II beta chain 2 (CK II beta 2 subunit).
 GN CKB2.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=11959847;
 RA Yang Y., Cheng P., Liu Y.;
 RT "Regulation of the Neurospora circadian clock by casein kinase II.";
 RL Genes Dev. 16:994-1006(2002).
 CC -!- FUNCTION: Plays a complex role in regulating the basal catalytic activity of the alpha subunit (By similarity).
 CC -!- SUBUNIT: Heterotetramer of 2 alpha subunits, a beta 1 subunit and a beta 2 subunit.
 CC -!- PTM: Phosphorylated by alpha chain (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE CASEIN KINASE 2 BETA CHAIN FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; AF494378; AAM14626.1; -;
 DR InterPro; IPR000704; CAS_kinase_II.
 DR Pfam; PF01214; CK II beta; 1.
 DR PRINTS; PR00472; CASKNINASEII.
 DR ProDom; PD003829; CAS kinase II; 1.
 DR PROSITE; PS01101; CK2_BETA; 1.
 KW Transferase; Serine/threonine-protein kinase; Phosphorylation.
 FT DOMAIN 221 259 GLU-RICH.
 SQ SEQUENCE 285 AA; 32573 MW; 3E9ADAB98530194 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 285;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 VPAKSV 16
 |||||

Db 179 VPAKSV 184

RESULT 12

ID COBB THEVO STANDARD; PRT; 454 AA.
 AC Q97BP4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable cobyrinic acid A,C-diamide synthase.
 GN COBB OR TV0411 OR TVG0399197.
 OS Thermoplasma volcanium.
 OC Archaea; Euryarchaeota; Thermoplasma; Thermoplasmatales;
 OC Thermoplasmataceae; Thermoplasma.
 OX NCBI_TaxID=50339;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GSS1 / DSM 4299 / JCM 9571;
 RX MEDLINE=20570466; PubMed=11121031;
 RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
 RA Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
 RA Nunoshiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
 RT "Archaeal adaptation to higher temperatures revealed by genomic sequence of Thermoplasma volcanium.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
 CC -!- FUNCTION: Responsible for the amidation of carboxylic groups at position A and C of either cobyrinic acid or hydrogenobrynic acid. NH(2) groups are provided by glutamine, and one molecule of ATP is hydrolyzed for each amidation (By similarity).
 CC -!- PATHWAY: Cbbalamin biosynthesis.
 CC -!- SIMILARITY: Belongs to the COBB/cobQ family. COBB subfamily.
 CC -----
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AP000992; BAB59553.1; -;
 DR HAMAP; MF 00027; -; 1.
 DR InterPro; IPR004484; CbiA.
 DR InterPro; IPR002586; CbiA_P.
 DR Pfam; PF01656; CbiA; 1.
 DR TIGRFAMs; TIGR00379; COBB; 1.
 KW Cobalamin biosynthesis; Complete proteome.
 SQ SEQUENCE 454 AA; 50354 MW; 8C2B1C0D821FF086 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 454;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GIVPAK 14
 |||||
 DB 349 GIVPAK 354

RESULT 13

ID SDHL ECOLI STANDARD; PRT; 454 AA.
 AC P16095;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE L-serine dehydratase 1 (EC 4.3.1.17) (L-serine deaminase 1) (SDH 1) (L-SD1).
 DE SDAA OR B1814.
 GN Escherichia coli.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;

RESULT 15

POLG_HCVJT STANDARD; PRT; 3010 AA.

AC Q00269;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P67); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

OS Hepatitis C virus (Isolate HC-JT) (HCV).

OC Viruses; asRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

OX NCBI_TaxID=31642;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92295714; PubMed=1318627;

RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J., .

RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;

RT "Molecular cloning of hepatitis C virus genome from a single Japanese carrier: sequence variation within the same individual and among infected individuals."

RT Virus Res. 23:39-53(1992).

RL

CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.

CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; D11168; BAA01943.1; -.

DR PIR; A45573; A45573.

DR PDB; 1AIQ; 25-MAR-98.

DR PDB; 1UXP; 14-JAN-98.

DR MEROPS; S29.001; -.

DR MEROPS; U39.001; -.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV capsid.

DR InterPro; IPR002521; HCV_core.

DR InterPro; IPR002519; HCV_env.

DR InterPro; IPR002531; HCV NS1.

DR InterPro; IPR002518; HCV NS2.

DR InterPro; IPR004109; HCV NS3.

DR InterPro; IPR000745; HCV NS4a.

DR InterPro; IPR001490; HCV NS4b.

DR InterPro; IPR002868; HCV NS5a.

DR InterPro; IPR002166; HCV RdRP.

DR InterPro; IPR007095; RNA pol DS PS.

DR InterPro; IPR007094; RNA pol PSvir.

DR Pfam; PF01543; HCV capsid; 1.

DR Pfam; PF01542; HCV_core; 1.

DR Pfam; PF01539; HCV_env; 1.

DR Pfam; PF01560; HCV_NS1; 1.

DR Pfam; PF01538; HCV_NS2; 1.

DR Pfam; PF02907; HCV_NS3; 1.

DR Pfam; PF01006; HCV_NS4a; 1.

DR Pfam; PF01001; HCV_NS4b; 1.

DR Pfam; PF01506; HCV_NS5a; 1.

DR Pfam; PF00271; helicase_C; 1.

DR Pfam; PF00998; Viral_RdRP; 1.

DR Pfam; PF018062; HCV_NS1; 1.

DR ProDom; PD186062; HCV_NS1; 1.

DR SMART; SM00487; DEXDC; 1.

KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;

KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;

KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;

KW 3D-structure.

FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE

FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.

FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).

FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).

FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).

FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).

FT CHAIN 1007 1615 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).

FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).

FT CHAIN 1869 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).

FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).

FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).

FT ACT_SITE 1083 1083 POTENTIAL.

FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT NP_BIND 1230 1237 ATP (POTENTIAL).

FT SITE 1316 1319 DECH BOX.

FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).

SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;

Query Match 30.0%; Score 6; DB 1; Length 3010;

Best Local Similarity 100.0%; Pred. No. 78;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 CGIIPA 13

Db 494 CGIIPA 499

Search completed: November 21, 2003, 20:59:56

Job time : 5.85 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:45:57 ; Search time 24.4 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-79

Perfect score: 20

Sequence: 1 WHYPPKPGIVPAKSVCGPV 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	20	100.0	64 12 Q8JPF6	Q8JPF6 hepatitis c
2	20	100.0	64 12 Q8JPF9	Q8JPF9 hepatitis c
3	20	100.0	148 12 Q8JPGA	Q8JPGA hepatitis c
4	20	100.0	151 12 Q9YK51	Q9YK51 hepatitis c
5	20	100.0	151 12 Q91987	Q91987 hepatitis c
6	20	100.0	151 12 Q91964	Q91964 hepatitis c
7	20	100.0	151 12 Q91966	Q91966 hepatitis c
8	20	100.0	151 12 Q9YK55	Q9YK55 hepatitis c
9	20	100.0	151 12 Q9YK60	Q9YK60 hepatitis c
10	20	100.0	151 12 Q9YK28	Q9YK28 hepatitis c
11	20	100.0	151 12 Q9YIG4	Q9YIG4 hepatitis c
12	20	100.0	151 12 Q91965	Q91965 hepatitis c
13	20	100.0	151 12 Q9YK25	Q9YK25 hepatitis c
14	20	100.0	151 12 Q9YJ79	Q9YJ79 hepatitis c
15	20	100.0	151 12 Q9YK61	Q9YK61 hepatitis c
16	20	100.0	151 12 Q9YK59	Q9YK59 hepatitis c

17	20	100.0	151 12 Q9YJ17	Q9YJ17 hepatitis c
18	20	100.0	151 12 Q91985	Q91985 hepatitis c
19	20	100.0	151 12 Q9YIH4	Q9YIH4 hepatitis c
20	20	100.0	151 12 Q9YK54	Q9YK54 hepatitis c
21	20	100.0	151 12 Q91963	Q91963 hepatitis c
22	20	100.0	151 12 Q9YK27	Q9YK27 hepatitis c
23	20	100.0	151 12 Q9YK58	Q9YK58 hepatitis c
24	20	100.0	151 12 Q9YK53	Q9YK53 hepatitis c
25	20	100.0	151 12 Q91988	Q91988 hepatitis c
26	20	100.0	151 12 Q9YJ92	Q9YJ92 hepatitis c
27	20	100.0	151 12 Q91986	Q91986 hepatitis c
28	20	100.0	151 12 Q9YK63	Q9YK63 hepatitis c
29	20	100.0	151 12 Q9YK62	Q9YK62 hepatitis c
30	20	100.0	162 12 Q8JPG1	Q8JPG1 hepatitis c
31	20	100.0	191 12 Q8JPL7	Q8JPL7 hepatitis c
32	20	100.0	191 12 Q8JPL6	Q8JPL6 hepatitis c
33	20	100.0	191 12 Q8JPL5	Q8JPL5 hepatitis c
34	20	100.0	191 12 Q8JPG3	Q8JPG3 hepatitis c
35	20	100.0	191 12 Q8JPG7	Q8JPG7 hepatitis c
36	20	100.0	191 12 Q8JPG5	Q8JPG5 hepatitis c
37	20	100.0	191 12 Q8BBT2	Q8BBT2 hepatitis c
38	20	100.0	191 12 Q8BBT1	Q8BBT1 hepatitis c
39	20	100.0	191 12 Q8BBT0	Q8BBT0 hepatitis c
40	20	100.0	191 12 Q8BBB9	Q8BBB9 hepatitis c
41	20	100.0	191 12 Q8BBB8	Q8BBB8 hepatitis c
42	20	100.0	191 12 Q8BBB6	Q8BBB6 hepatitis c
43	20	100.0	191 12 Q8BBB3	Q8BBB3 hepatitis c
44	20	100.0	191 12 Q8BBB1	Q8BBB1 hepatitis c
45	20	100.0	191 12 Q8BBB8	Q8BBB8 hepatitis c

ALIGNMENTS

RESULT 1 0 .
Q8JPF6 PRELIMINARY; PRT; 64 AA.
ID Q8JPF6;
AC Q8JPF6;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DE 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JF3C9;
RA Wang X.-P., Gavel C., Cunha C.B., Laser B.S., Sahn B.A., Kahn P.D.,
RA Kaplan M.H.;
RT "The Genetic Diversity of Hepatitis C Virus in HIV-1 Co-infected Patients";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF529759; AAM96110.1;
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; I.
DR Pfam; PF01560; HCV NS1; I.
DR ProDom; PD186062; HCV NS1; I.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 64
SQ SEQUENCE 64 AA; 7061 MW; 385E456BAA83020A CRC64;
Query Match 100.0%; Score 20; DB 12; Length 64;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WHYPPKPGIVPAKSVCGPV 20
DB 48 WHYPPKPGIVPAKSVCGPV 62

```

RESULT 2
QBPF9
ID Q8JPF9 PRELIMINARY; PRT; 64 AA.
AC Q8JPF9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JF3C6;
RA Wang X.-P., Gavel C., Cunha C.B., Laser B.S., Sahn B.A., Kahn P.D.,
RA Kaplan M.H.;
RT "The Genetic Diversity of Hepatitis C Virus in HIV-1 Co-infected
RT Patients.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF529756; AAM96107.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 64
FT NON_TER 64
SQ SEQUENCE 64 AA; 7061 MW; 395E456BA83020A CRC64;

Query Match 100.0%; Score 20; DB 12; Length 64;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPGIVPAKSVCGPV 20
DB 43 WHYPPKPGIVPAKSVCGPV 62

RESULT 3
QBPG4
ID Q8JPG4 PRELIMINARY; PRT; 148 AA.
AC Q8JPG4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JF2C16;
RA Wang X.-P., Gavel C., Cunha C.B., Laser B.S., Sahn B.A., Kahn P.D.,
RA Kaplan M.H.;
RT "The Genetic Diversity of Hepatitis C Virus in HIV-1 Co-infected
RT Patients.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF529751; AAM96102.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 2.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 148
FT NON_TER 148

```

```

SQ SEQUENCE 148 AA; 15910 MW; FB241847B48B59DF CRC64;

Query Match 100.0%; Score 20; DB 12; Length 148;
Best Local Similarity 100.0%; Pred. No. 3.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPGIVPAKSVCGPV 20
DB 127 WHYPPKPGIVPAKSVCGPV 146

RESULT 4
Q9YK51
ID Q9YK51 PRELIMINARY; PRT; 151 AA.
AC Q9YK51;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411414; PubMed=9738054;
RA Wang Y.M., Ray S.C., Laeyendecker O., Ticehurst J.R., Thomas D.L.;
RT "Assessment of hepatitis C virus sequence complexity by
RT electrophoretic mobilities of both single- and double-stranded DNAs.";
RL J. Clin. Microbiol. 36:2982-2989(1998).
DR EMBL; AF073098; AAC61398.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT CHAIN 1 >22 E1.
FT CHAIN 23 >151 E2.
FT NON_TER 151
FT NON_TER 151
SQ SEQUENCE 151 AA; 16134 MW; 3834D8F453019B82 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 151;
Best Local Similarity 100.0%; Pred. No. 3.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPGIVPAKSVCGPV 20
DB 126 WHYPPKPGIVPAKSVCGPV 145

RESULT 5
Q91987
ID Q91987 PRELIMINARY; PRT; 151 AA.
AC Q91987;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411414; PubMed=9738054;
RA Wang Y.M., Ray S.C., Laeyendecker O., Ticehurst J.R., Thomas D.L.;
RT "Assessment of hepatitis C virus sequence complexity by
RT electrophoretic mobilities of both single- and double-stranded DNAs.";
RL J. Clin. Microbiol. 36:2982-2989(1998).
DR EMBL; AF073144; AAC61444.1; -.

```



```

DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT CHAIN 1 1
FT CHAIN 23 >22 E1.
FT CHAIN 23 >151 E2.
FT NON_TER 151 151
SQ SEQUENCE 151 AA; 16309 MW; 5AB3DB60B6877DB6 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 151;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPGGIVPAKSVCGPV 20
DB 126 WHYPPKPGGIVPAKSVCGPV 145

RESULT 6
ID O91964 PRELIMINARY; PRT; 151 AA.
AC O91964;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411414; PubMed=9738054;
RA Wang Y.M., Ray S.C., Laeyendecker O., Ticehurst J.R., Thomas D.L.;
RT "Assessment of hepatitis C virus sequence complexity by
RT electrophoretic mobilities of both single- and double-stranded DNAs.";
RL J. Clin. Microbiol. 36:2982-2989(1998).
DR EMBL; AF073077; AAC61377.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT CHAIN 1 1
FT CHAIN 23 >22 E1.
FT CHAIN 23 >151 E2.
FT NON_TER 151 151
SQ SEQUENCE 151 AA; 16279 MW; F2804C4B2A99FC2A CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 151;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPGGIVPAKSVCGPV 20
DB 126 WHYPPKPGGIVPAKSVCGPV 145

RESULT 7
ID O91966 PRELIMINARY; PRT; 151 AA.
AC O91966;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (fragment).
OS Hepatitis C virus.

```

```

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411414; PubMed=9738054;
RA Wang Y.M., Ray S.C., Laeyendecker O., Ticehurst J.R., Thomas D.L.;
RT "Assessment of hepatitis C virus sequence complexity by
RT electrophoretic mobilities of both single- and double-stranded DNAs.";
RL J. Clin. Microbiol. 36:2982-2989(1998).
DR EMBL; AF073086; AAC61386.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT CHAIN 1 1
FT CHAIN 23 >22 E1.
FT CHAIN 23 >151 E2.
FT NON_TER 151 151
SQ SEQUENCE 151 AA; 16004 MW; 66C430F04EE81F20 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 151;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPGGIVPAKSVCGPV 20
DB 126 WHYPPKPGGIVPAKSVCGPV 145

RESULT 8
ID O9YK55 PRELIMINARY; PRT; 151 AA.
AC O9YK55;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411414; PubMed=9738054;
RA Wang Y.M., Ray S.C., Laeyendecker O., Ticehurst J.R., Thomas D.L.;
RT "Assessment of hepatitis C virus sequence complexity by
RT electrophoretic mobilities of both single- and double-stranded DNAs.";
RL J. Clin. Microbiol. 36:2982-2989(1998).
DR EMBL; AF073094; AAC61394.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT CHAIN 1 1
FT CHAIN 23 >22 E1.
FT CHAIN 23 >151 E2.
FT NON_TER 151 151
SQ SEQUENCE 151 AA; 16128 MW; 60F2722E56022315 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 151;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPGGIVPAKSVCGPV 20
DB 126 WHYPPKPGGIVPAKSVCGPV 145

```

```
RESULT 9
Q9YK60
ID Q9YK60 PRELIMINARY; PRT; 151 AA.
AC Q9YK60;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411414; PubMed=9738054;
RA Wang Y.M., Ray S.C., Laeyendecker O., Ticehurst J.R., Thomas D.L.;
RT "Assessment of hepatitis C virus sequence complexity by
RT electrophoretic mobilities of both single- and double-stranded DNAs.";
RL J. Clin. Microbiol. 36:2982-2989(1998).
DR EMBL; AF073085; AAC61385.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT CHAIN 1 >22 E1.
FT NON_TER 1
FT CHAIN 23 >151 E2.
FT NON_TER 151
FT SEQUENCE 151 AA; 16142 MW; C4CEP048918CQF24 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 151;
Best Local Similarity 100.0%; Pred. No. 3.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPGIVPAKSVCGPV 20
Db 126 WHYPPKPGIVPAKSVCGPV 145

RESULT 10
Q9YK28
ID Q9YK28 PRELIMINARY; PRT; 151 AA.
AC Q9YK28;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411414; PubMed=9738054;
RA Wang Y.M., Ray S.C., Laeyendecker O., Ticehurst J.R., Thomas D.L.;
RT "Assessment of hepatitis C virus sequence complexity by
RT electrophoretic mobilities of both single- and double-stranded DNAs.";
RL J. Clin. Microbiol. 36:2982-2989(1998).
DR EMBL; AF073146; AAC61446.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT CHAIN 1 >22 E1.
FT NON_TER 1
```

```
FT CHAIN 23 >151 E2.
FT NON_TER 151
FT SEQUENCE 151 AA; 16210 MW; 56C2C710E68761C6 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 151;
Best Local Similarity 100.0%; Pred. No. 3.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPGIVPAKSVCGPV 20
Db 126 WHYPPKPGIVPAKSVCGPV 145

RESULT 11
Q9YIG4
ID Q9YIG4 PRELIMINARY; PRT; 151 AA.
AC Q9YIG4;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411414; PubMed=9738054;
RA Wang Y.M., Ray S.C., Laeyendecker O., Ticehurst J.R., Thomas D.L.;
RT "Assessment of hepatitis C virus sequence complexity by
RT electrophoretic mobilities of both single- and double-stranded DNAs.";
RL J. Clin. Microbiol. 36:2982-2989(1998).
DR EMBL; AF073073; AAC61373.1; -.
DR EMBL; AF073072; AAC61372.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT CHAIN 1 >22 E1.
FT NON_TER 1
FT CHAIN 23 >151 E2.
FT NON_TER 151
FT SEQUENCE 151 AA; 16208 MW; CCFF7C43B0C6C621 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 151;
Best Local Similarity 100.0%; Pred. No. 3.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPPKPGIVPAKSVCGPV 20
Db 126 WHYPPKPGIVPAKSVCGPV 145

RESULT 12
Q91965
ID Q91965 PRELIMINARY; PRT; 151 AA.
AC Q91965;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411414; PubMed=9738054;
RA Wang Y.M., Ray S.C., Laeyendecker O., Ticehurst J.R., Thomas D.L.;
RT "Assessment of hepatitis C virus sequence complexity by
```

```

RT electrophoretic mobilities of both single- and double-stranded DNAs.";
RL J. Clin. Microbiol. 36:2982-2989(1998).
DR EMBL; AF073082; AAC61382.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT CHAIN 1 >22 E1.
FT CHAIN 23 >151 E2.
FT NON_TER 151
SQ SEQUENCE 151 AA; 16118 MW; 12174F2E47576407 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 151;
Best Local Similarity 100.0%; Pred. No. 3.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPKPCGIVPAKSVCGPV 20
Db 126 WHYPKPCGIVPAKSVCGPV 145

RESULT 13
Q9YK25 PRELIMINARY; PRT; 151 AA.
AC Q9YK25;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411414; PubMed=9738054;
RA Wang Y.M., Ray S.C., Laeyendecker O., Ticehurst J.R., Thomas D.L.;
RT "Assessment of hepatitis C virus sequence complexity by
RT electrophoretic mobilities of both single- and double-stranded DNAs.";
RL J. Clin. Microbiol. 36:2982-2989(1998).
DR EMBL; AF073150; AAC61450.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT CHAIN 1 >22 E1.
FT CHAIN 23 >151 E2.
FT NON_TER 151
SQ SEQUENCE 151 AA; 16281 MW; 04A3CD070F61AA0 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 151;
Best Local Similarity 100.0%; Pred. No. 3.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPKPCGIVPAKSVCGPV 20
Db 126 WHYPKPCGIVPAKSVCGPV 145

RESULT 14
Q9YJ79 PRELIMINARY; PRT; 151 AA.
AC Q9YJ79;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)

```

```

DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411414; PubMed=9738054;
RA Wang Y.M., Ray S.C., Laeyendecker O., Ticehurst J.R., Thomas D.L.;
RT "Assessment of hepatitis C virus sequence complexity by
RT electrophoretic mobilities of both single- and double-stranded DNAs.";
RL J. Clin. Microbiol. 36:2982-2989(1998).
DR EMBL; AF073079; AAC61379.1; -.
DR EMBL; AF073078; AAC61378.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT CHAIN 1 >22 E1.
FT CHAIN 23 >151 E2.
FT NON_TER 151
SQ SEQUENCE 151 AA; 16149 MW; 8175DFDE146752F CRC64;

Query Match 100.0%; Score 20; DB 12; Length 151;
Best Local Similarity 100.0%; Pred. No. 3.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WHYPKPCGIVPAKSVCGPV 20
Db 126 WHYPKPCGIVPAKSVCGPV 145

RESULT 15
Q9YK61 PRELIMINARY; PRT; 151 AA.
AC Q9YK61;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411414; PubMed=9738054;
RA Wang Y.M., Ray S.C., Laeyendecker O., Ticehurst J.R., Thomas D.L.;
RT "Assessment of hepatitis C virus sequence complexity by
RT electrophoretic mobilities of both single- and double-stranded DNAs.";
RL J. Clin. Microbiol. 36:2982-2989(1998).
DR EMBL; AF073080; AAC61380.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT CHAIN 1 >22 E1.
FT CHAIN 23 >151 E2.
FT NON_TER 151
SQ SEQUENCE 151 AA; 16147 MW; 5A9505F42749F416 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 151;
Best Local Similarity 100.0%; Pred. No. 3.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 WHYPKPCGIVPAKSVCGPV 20
| | | | | | | | | | | | | | | | | |
Db 126 WHYPKPCGIVPAKSVCGPV 145

Search completed: November 21, 2003, 21:08:17
Job time : 25.45 secs

0 .

0 .

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:22 ; Search time 32.1 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-80

Perfect score: 20

Sequence: 1 AKSVCVPYCFPTSPVVGVT 20

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	20	100.0	20 16 AAR84509	Hepatitis C virus
2	20	100.0	20 17 AAR91003	HCV E2 peptide E2-
3	20	100.0	20 23 AAO18709	Hepatitis C virus
4	20	100.0	30 23 AAU84631	HCV HepC1a segment
5	20	100.0	43 23 ABB77236	Selected interacti
6	20	100.0	96 23 ABB77258	HCV bait polypepti
7	20	100.0	176 23 ABB77257	HCV bait polypepti
8	20	100.0	192 19 AAU67009	HCV nucleocapsid c
9	20	100.0	250 23 AAU79220	Hepatitis C Virus

10	20	100.0	254	22	AAB68043	Amino acid sequenc
11	20	100.0	271	22	AAE00445	HCV E2-634 HVR1 CO
12	20	100.0	271	22	AAE00446	HCV E2-634 HVR1-mu
13	20	100.0	278	14	AAR33997	Th E2/NS1 protein.
14	20	100.0	305	18	AAW00929	Recombinant HCV E2
15	20	100.0	333	14	AAR40118	HGH-HCV-E2 fusion
16	20	100.0	337	16	AAR79217	pHCV351-encoded AP
17	20	100.0	350	23	AAE19891	Hepatitis C virus
18	20	100.0	363	22	AAB68042	Amino acid sequenc
19	20	100.0	363	24	AAE32880	Hepatitis C virus
20	20	100.0	363	24	ABP55567	Hepatitis C virus
21	20	100.0	367	14	AAR40115	APP-HCV-E2 fusion
22	20	100.0	367	16	AAR79218	pHCV167-encoded pr
23	20	100.0	377	16	AAR79226	pHCV422-encoded AP
24	20	100.0	397	16	AAR79220	pHCV419-encoded AP
25	20	100.0	399	14	AAR40117	HGH-HCV-E2 fusion
26	20	100.0	402	14	AAR34438	Sequence of glycop
27	20	100.0	402	14	AAR34439	Sequence of glycop
28	20	100.0	402	14	AAR34440	Sequence of glycop
29	20	100.0	409	14	AAR33995	H77 E2/NS1 protein
30	20	100.0	409	14	AAR33996	H90 E2/NS1 protein
31	20	100.0	410	16	AAR79227	pHCV423-encoded AP
32	20	100.0	417	16	AAR79228	pHCV424-encoded AP
33	20	100.0	434	16	AAR79219	pHCV418-encoded AP
34	20	100.0	441	16	AAR79220	pHCV429-encoded AP
35	20	100.0	447	16	AAR79229	pHCV425-encoded AP
36	20	100.0	453	16	AAR79225	pHCV421-encoded AP
37	20	100.0	454	10	AAP90183	Sequence of hepati
38	20	100.0	454	10	AAP92049	Sequence encoded b
39	20	100.0	454	21	AAE18526	Protein encoded by
40	20	100.0	463	14	AAR33588	HCV CKS-NS1s1 fusi
41	20	100.0	463	14	AAR33187	Sequence of subfra
42	20	100.0	463	22	AAE69007	HCV recombinant an
43	20	100.0	480	14	AAR33992	HCV-1 E2/NS1 prote
44	20	100.0	490	15	AAR79224	pHCV420-encoded AP
45	20	100.0	497	22	AAE00443	HCV HVR1-7 constru

ALIGNMENTS

RESULT 1

AAR84509

ID AAR84509 standard; peptide; 20 AA.

XX

AC AAR84509;

XX

DT 06-JAN-Q997 (first entry)

XX

DE Hepatitis C virus peptide NS1-9 (residues 499-518).

XX

DE Hepatitis C virus; HCV; immunogen; non-structural region; NS1;

XX immunodominant; T cell epitope; vaccine.

KW

OS Hepatitis C virus.

XX

PN WO9512677-A2.

XX

PD 11-MAY-1995.

XX

PF 28-OCT-1994; 94WO-EP03555.

XX

PR 04-NOV-1993; 93EP-0402718.

XX

PA (INNO-) INNOGENETICS NV.

XX

PI Deleys R, Leroux-Roels G, Maertens G;

XX

DR WPI; 1995-193822/25.

XX

PT Hepatitis C virus immunogenic polypeptide contg. a T-cell

PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in

PT production of vaccines, therapeutic agents, etc.

XX Example 4; Page 51; 105pp; English.

PS A series of overlapping peptides (including the present sequence) was

CC synthesised based on sequences in the core, E1 and E2/NS1 regions of

CC hepatitis C virus. The peptides were used as antigens in lympho-

CC proliferative assays to identify the main T-cell epitopes.

XX Sequence 20 AA;

SQ Query Match 100.0%; Score 20; DB 16; Length 20;

Best Local Similarity 100.0%; Pred. No. 4.7e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20

DB 1 AKSVCGPVYCFPTSPVVVGT 20

RESULT 2

AA018709

ID AAR91003 standard; peptide; 20 AA.

XX AC AAR91003;

XX DT 25-SEP-1996 (first entry)

XX DE HCV E2 peptide E2-9B for competition studies.

XX HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;

KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.

KW Synthetic.

OS WO9604385-A2.

XX PN 15-FEB-1996.

XX PD 31-JUL-1995; 95WO-EP03031.

XX PF 29-JUL-1994; 94EP-0870132.

XX PR (INNO-) INNOGENETICS NV.

XX PA Bosman F, Buyse M, De Martynoff G, Maertens G;

XX PI WPI; 1996-129401/13.

XX DR Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope

XX PT proteins - in presence of disulphide bond cleavage agent, to

PT produce proteins suitable for direct use in vaccines or diagnostic

PT assays of HCV

XX Example 7; Page 67; 146pp; English.

PS AAR90965-R90980 and AAR90995-R91015 represent synthetic hepatitis C

XX virus (HCV) E1 and E2 peptides used in competition studies. This

CC sequence represents a synthetic E2 peptide, and corresponds to residues

CC 499-518 of the E2 protein sequence. These sequences are useful for in

CC vitro monitoring of HCV disease, or prognosis of the response to

CC interferon treatment of patients suffering from HCV infection. These

CC sequences compete with the proteins produced by AAT12704-T12709 and

CC AAT12961-T12974, which are included in vectors for the production of

CC recombinant E1, E2, and E1/E2. The recombinant proteins can then be

CC isolated and purified by carrying out a disulphide bond cleavage, or a

CC reduction step with a disulphide bond cleavage agent, after lysis of

CC recombinant host cells. The constructs containing the purified HCV

CC envelope proteins can be used for vaccinating humans against HCV, for in

CC vitro detection of HCV antibodies in a sample, and in a serotyping assay

CC for detecting one or more serological types of HCV present in a

CC biological sample. The constructs can also be immobilised on a solid

CC substrate and incorporated into a reversed phase hybridisation assay for

CC determining the presence or the genotype of HCV. The new purification

method preserves the conformation of the recombinantly expressed E1, E2

and E1/E2, and eliminates contaminating proteins. Antigens isolated

using this method are more reactive with human sera than those isolated

by known techniques.

XX Sequence 20 AA;

SQ Query Match 100.0%; Score 20; DB 17; Length 20;

Best Local Similarity 100.0%; Pred. No. 4.7e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20

DB 1 AKSVCGPVYCFPTSPVVVGT 20

RESULT 3

AA018709

ID AAO18709 standard; Peptide; 20 AA.

XX AC AAO18709;

XX DT 24-OCT-2002 (first entry)

XX DE Hepatitis C virus E2 protein derived peptide E2-9B.

XX HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;

KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.

KW Synthetic.

OS WO200255548-A2.

XX PN 18-JUL-2002.

XX PD 11-JAN-2002; 2002WO-EP00219.

XX PF 11-JAN-2001; 2001US-260699P.

XX PR 30-AUG-2001; 2001US-315768P.

XX PA (INNO-) INNOGENETICS NV.

XX PI Maertens G, Bosman F, Buyse M;

XX DR WPI; 2002-599657/64.

XX PT New therapeutic vaccine compositions comprising at least one purified

PT recombinant hepatitis C virus (HCV) single or specific oligomeric

PT recombinant envelope protein E1 or E2, useful for immunizing humans

XX from HCV infection

PS Example 7; Page 228; 243pp; English.

XX The present invention relates to new therapeutic vaccine compositions for

XX inducing hepatitis C virus (HCV)-specific antibodies, comprising a

CC composition containing at least one purified recombinant HCV single or

CC specific oligomeric recombinant envelope proteins selected from an E1 and

CC an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are

CC useful for inducing HCV-specific antibodies or for immunising humans

CC against HCV. The recombinant HCV E1 and/or E2 proteins are useful as

CC vaccines or therapeutics, in HCV screening and confirmatory antibody

CC tests, for raising antibodies, in the preparation of medicament, and for

CC in vitro monitoring of HCV disease or prognosing the response to

CC treatment of patients suffering from HCV infection. The present sequence

CC is a peptide derived from the proteins of the invention.

XX Sequence 20 AA;

SQ Query Match 100.0%; Score 20; DB 23; Length 20;

Best Local Similarity 100.0%; Pred. No. 4.7e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20

Matches	20;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	----	--------	----	------	----

RESULT 6
ABB77258

ID ABB77258 standard; Protein; 96 AA.
AC ABB77258;
XX
DT 28-JUN-2002 (first entry)
XX
DE HCV bait polypeptide 6.
XX
KW SID; selected interacting domain; HCV; hepatitis C virus;
KW liver disease; liver cancer; virucide; hepatotropic; antiinflammatory;
KW antibacterial.
XX
OS Hepatitis C virus strain H77.
XX
FN EPI178116-A1.
XX
PD 06-FEB-2002.
XX
PF 03-AUG-2000; 2000EP-0402225.
PR 03-AUG-2000; 2000EP-0402225.
XX
PA (HYBR-) HYBRIGENICS SA.
XX
PI Legrain P, Whiteside S, Wojcik J;
XX
DR WPI; 2002-208115/27.
DR N-PSDB; ABL55590.
XX
PT New selected interacting domain polypeptides and polynucleotides,
PT useful for treating or preventing infections or pathologies caused by
PT hepatitis C virus (HCV) or those linked to HCV infection -
XX
XX Claim 26; SEQ ID 82; 61pp + sequence listing; English.
PS
XX The invention relates to nucleic acids encoding polypeptides which are
CC termed SID polypeptides (selected interacting domain). These polypeptides
CC are the final products of a double selection method involving a first
CC step of selection of Hepatitis C virus (HCV)-derived polynucleotides
CC through a two-hybrid system, and a second selection step involving an
CC alignment between the different polynucleotides selected at the first
CC step. The activity of polypeptides of the invention may be described as,
CC virucide, hepatotropic, antiinflammatory and antibacterial. The
CC polypeptide, polynucleotide and compositions comprising them are useful
CC for treating or preventing viral or a bacterial infection, specifically
CC infections or pathologies caused by HCV, or those pathologies linked to
CC HCV infection. These may include liver disease and liver cancer. The
CC current sequence represents a HCV bait polypeptide.
CC Note: The sequence data for this patent is not represented in the
CC specification, but is based on sequence information supplied by the
CC European Patent Office.
XX
SQ Sequence 96 AA;
Query Match 100.0%; Score 20; DB 23; Length 96;
Best Local Similarity 100.0%; Pred. No. 1.6e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AKSVCGPVYCFPTSPVWVGT 20
Db 13 AKSVCGPVYCFPTSPVWVGT 32
RESULT 7
ABB77257
ID ABB77257 standard; Protein; 176 AA.
XX
AC ABB77257;
XX
DT 28-JUN-2002 (first entry)
XX
DE HCV bait polypeptide 5.
XX

KW SID; selected interacting domain; HCV; hepatitis C virus;
KW liver disease; liver cancer; virucide; hepatotropic; antiinflammatory;
KW antibacterial.
XX
OS Hepatitis C virus strain H77.
XX
FN EPI178116-A1.
XX
PD 06-FEB-2002.
XX
PF 03-AUG-2000; 2000EP-0402225.
PR 03-AUG-2000; 2000EP-0402225.
XX
PA (HYBR-) HYBRIGENICS SA.
XX
PI Legrain P, Whiteside S, Wojcik J;
XX
DR WPI; 2002-208115/27.
DR N-PSDB; ABL55589.
XX
PT New selected interacting domain polypeptides and polynucleotides,
PT useful for treating or preventing infections or pathologies caused by
PT hepatitis C virus (HCV) or those linked to HCV infection -
XX
XX Claim 26; SEQ ID 81; 61pp + sequence listing; English.
PS
XX The invention relates to nucleic acids encoding polypeptides which are
CC termed SID polypeptides (selected interacting domain). These polypeptides
CC are the final products of a double selection method involving a first
CC step of selection of Hepatitis C virus (HCV)-derived polynucleotides
CC through a two-hybrid system, and a second selection step involving an
CC alignment between the different polynucleotides selected at the first
CC step. The activity of polypeptides of the invention may be described as,
CC virucide, hepatotropic, antiinflammatory and antibacterial. The
CC polypeptide, polynucleotide and compositions comprising them are useful
CC for treating or preventing viral or a bacterial infection, specifically
CC infections or pathologies caused by HCV, or those pathologies linked to
CC HCV infection. These may include liver disease and liver cancer. The
CC current sequence represents a HCV bait polypeptide.
CC Note: The sequence data for this patent is not represented in the
CC specification, but is based on sequence information supplied by the
CC European Patent Office.
XX
SQ Sequence 176 AA;
Query Match 100.0%; Score 20; DB 23; Length 176;
Best Local Similarity 100.0%; Pred. No. 2.6e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AKSVCGPVYCFPTSPVWVGT 20
Db 31 AKSVCGPVYCFPTSPVWVGT 50
RESULT 8
AAW67009
ID AAW67009 standard; protein; 192 AA.
XX
AC AAW67009;
XX
DT 02-MAR-1999 (first entry)
XX
DE HCV nucleocapsid core protein.
XX
KW Hepatitis C virus; HCV; nucleocapsid; core protein; analogue; antibody;
KW non-structural protein; thioamide bond; peptide bond.
XX
OS Hepatitis C virus.
XX
FH Key Location/Qualifiers
FT Misc-difference 79 /note= "given in specification as Try"

XX JP10226698-A.
 XX
 XX 25-AUG-1998.
 XX
 XX 19-FEB-1997; 97JP-0034702.
 XX
 XX 19-FEB-1997; 97JP-0034702.
 XX
 XX (KYOWA) KYOWA MEDEX KK.
 XX
 XX WPI; 1998-515103/44.
 XX
 XX Determination of antibody in sample - uses peptide analog absorbed
 PT or chemically bound on carrier as antigen
 XX
 XX Disclosure; Page 4; 13pp; Japanese.
 XX
 XX This sequence represents the Hepatitis C virus (HCV) nucleocapsid core
 CC protein. The invention relates to peptide analogues derived from HCV
 CC proteins, e.g. AA67417-W67426, which can be used for the determination
 CC of anti-HCV antibodies in a sample. Preferably the peptide analogues
 CC contain one or more thioamide peptide bonds where at least one oxygen
 CC atom of the peptide bond is replaced by sulphur atom. The peptide
 CC analogues can be adsorbed or chemically bound to a carrier.
 XX
 XX Sequence 192 AA;
 SQ
 Query Match 100.0%; Score 20; DB 19; Length 192;
 Best Local Similarity 100.0%; Pred. No. 2.7e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AKSVCGPVYCFPTSPVWVG 20
 DB 49 AKSVCGPVYCFPTSPVWVG 68
 RESULT 9
 AAU79220
 ID AAU79220 standard; Protein; 250 AA.
 XX
 XX AAU79220;
 XX
 XX 15-JUL-2002 (first entry)
 DT
 DE Hepatitis C Virus (HCV) delta-delta E2 genotype 1a protein.
 XX
 XX Hepatitis C Virus E2; HCV E2; virucide; hepatotropic; IL-2;
 KW antiinflammatory; HCV infection; interleukin-2; gamma-interferon;
 KW granulocyte macrophage-colony stimulating factor; GM-CSF;
 KW delta-delta E2 genotype 1a.
 XX
 XX Hepatitis C Virus.
 OS
 XX WO200222155-A1.
 PN
 XX 21-MAR-2002.
 PD
 XX
 XX 13-SEP-2001; 2001WO-US28767.
 PF
 XX 13-SEP-2000; 2000US-230927P.
 PR
 XX (HAWAII) HAWAII BIOTECHNOLOGY GROUP INC.
 PA
 XX Nakano ET, Clements DE, Humphreys T;
 PI
 XX WPI; 2002-383102/41.
 DR
 XX N-PSDB; ABK49386.
 DR
 XX New immunogenic polypeptide comprising hepatitis C virus E2 polypeptide
 PT useful for treating hepatitis C virus infection and for providing
 PT immune protection against virus infection
 XX

PS Claim 4; Fig 6; 84pp; English.
 XX
 XX The invention relates to a secreted polypeptide comprising hepatitis C
 CC virus (HCV) E2 polypeptide lacking all or a portion of its membrane
 CC spanning domain so that the E2 polypeptide is capable of secretion into
 CC growth medium when expressed recombinantly in a host cell. The
 CC polypeptide may also lack a portion of its C-terminus. The HCV E2
 CC secreted polypeptide is useful for producing anti-HCV antibodies. A
 CC purified immunogenic polypeptide comprising HCV E2 is useful for treating
 CC HCV infection and for providing immune protection against HCV infection
 CC by administering it to a subject having or at risk of having HCV
 CC infection or in need of protection. The method further comprises
 CC administering an immunomodulatory agent such as interleukin-2 (IL-2),
 CC granulocyte macrophage-colony stimulating factor (GM-CSF) or
 CC gamma-interferon. The polypeptide is useful as a vaccine, and with other
 CC HCV proteins to form a multi-component HCV vaccine for prophylactic or
 CC therapeutic treatment of HCV infection. This sequence represents an N-
 CC and C-terminally truncated HCV delta-delta E2 genotype 1a protein.
 XX
 XX Sequence 250 AA;
 SQ
 Query Match 100.0%; Score 20; DB 23; Length 250;
 Best Local Similarity 100.0%; Pred. No. 3.4e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AKSVCGPVYCFPTSPVWVG 20
 DB 88 AKSVCGPVYCFPTSPVWVG 107
 RESULT 10
 AAB68043
 ID AAB68043 standard; protein; 254 AA.
 XX
 XX AAB68043;
 AC
 XX 29-JUN-2001 (first entry)
 DT
 DE Amino acid sequence of water soluble variant of envelope E2 protein.
 XX
 XX Eo protein; HCV; envelope protein; E2 protein; HCV infection;
 KW HCV attachment.
 KW
 XX Synthetic.
 OS
 XX Hepatitis C virus.
 XX
 XX WO200122984-A1.
 PN
 XX 05-APR-2001.
 PD
 XX 26-SEP-2000; 2000WO-US26395.
 PF
 XX 29-SEP-1999; 99US-0407430.
 PR
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 PA
 XX Worman HJ, Mamiya N;
 PI
 XX WPI; 2001-273486/28.
 DR
 XX Treating or preventing hepatitis C virus infection in a subject,
 PT involves administering hepatitis C virus envelope protein E2 binding
 PT agents
 XX
 XX Claim 5; Fig 8; 46pp; English.
 PS
 XX The present sequence represents a water soluble variant of a Hepatitis C
 CC virus (HCV) envelope E2 protein. Agents that bind to the HCV E2 protein
 CC (such as the human Eo protein), and so inhibit the attachment of HCV onto
 CC cells (especially liver cells), are used to treat HCV infections in
 CC mammals, in particular humans. The specification also describes a method
 CC for identifying a compound which can be used for treating or preventing
 CC HCV in a subject and which can inhibit the attachment of HCV onto cells

CC by inhibiting the binding of HCV envelope E2 protein to a cellular
 CC protein associated with HCV attachment and entry into cells. The method
 CC comprises incubating the compound, HCV envelope E2 protein or its variant
 CC and a cellular protein capable of specifically binding to the HCV E2
 CC protein under suitable reaction conditions; determining the interactions
 CC between HCV envelope E2 protein and cellular protein in the presence and
 CC absence of the compound; and comparing the interaction to identify a
 CC compound which can inhibit the attachment of HCV onto cells.

XX Sequence 254 AA;

Query Match 100.0%; Score 20; DB 22; Length 254;
 Best Local Similarity 100.0%; Pred. No. 3.4e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFTPSPVVVGT 20
 |||||
 Db 94 AKSVCGPVYCFTPSPVVVGT 113

RESULT 11

AAE00445
 ID AAE00445 standard; Protein; 271 AA.

AC AAE00445;

DT 19-JUN-2001 (first entry)

XX HCV E2-634 HVR1 construct containing E1 signal sequence and truncated E2.
 DE Hepatitis C virus; HCV; hypervariable region one; HVR1; vaccine;
 KW antiviral; gene therapy; envelope 2 protein; E2; immunisation;
 KW HCV infection; viral replication; passive immunoprophylaxis.

XX Hepatitis C virus.

OS Synthetic.

XX Key Location/Qualifiers
 FH 1..20
 FT Peptide /label= Signal_peptide

FT /note= "Derived from endoplasmic reticulum E1
 signal sequence (364-383 amino acids)"

FT Protein 21..271

FT /label= Mature_C_terminal_truncated_E2_protein

XX WO200121807-A1.

PN 29-MAR-2001.

XX 22-SEP-2000; 2000WO-US25987.

XX 23-SEP-1999; 99US-0155823.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Fornis X, Bukh J, Emerson SU, Purcell RH;

XX WPI; 2001-266076/27.

XX Novel nucleic acid molecules that encode hepatitis C virus envelope 2
 PT protein lacking all or part of hypervariable region 1 of envelope
 PT protein, useful as vaccine components for treating or preventing HCV
 PT infections -

XX Example; Page -: 80pp; English.

XX The present sequence is hepatitis C virus (HCV) E2-634 HVR1 construct
 CC containing the endoplasmic reticulum signal sequence of envelope protein
 CC E1 and carboxy-terminal truncated E2 protein lacking the hypervariable
 CC region one (HVR1). The HCV E2 protein lacking HVR1 DNA is useful for
 CC producing infectious HCV and chimeric HCV viruses which are useful
 CC for identifying cell lines capable of supporting the replication of
 CC viruses. The infectious HCV and HVR1-chimeric HCV are used in the

CC production of attenuated or inactivated vaccines which are useful for
 CC treating or preventing HCV in a mammal by immunisation. The host cells
 CC expressing the H77C(HVR1) DNA is useful as an immunogen to stimulate a
 CC protective immune response to HCV. The immunogens are useful for
 CC producing protective antibodies to HCV. The antibodies produced are used
 CC in passive immunoprophylaxis for treatment of diseases caused by HCV in
 CC animals, especially humans. The H77C(HVR1) DNA is also useful in gene
 CC therapy.
 CC Note: The present sequence is not shown in the specification but is
 CC derived from Hepatitis C virus envelope 2 protein lacking hypervariable
 CC region 1 (AAE00442) referred as SEQ ID NO: 2 and shown in figure 1.

XX Sequence 271 AA;

Query Match 100.0%; Score 20; DB 22; Length 271;
 Best Local Similarity 100.0%; Pred. No. 3.6e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFTPSPVVVGT 20
 |||||
 Db 109 AKSVCGPVYCFTPSPVVVGT 128

RESULT 12

AAE00446
 ID AAE00446 standard; Protein; 271 AA.

XX AAE00446;

XX 19-JUN-2001 (first entry)

XX HCV E2-634 HVR1-mut containing E1 signal sequence and truncated E2.

DE Hepatitis C virus; HCV; hypervariable region one; HVR1; vaccine;

KW antiviral; gene therapy; envelope 2 protein; E2; immunisation; mutein;

KW HCV infection; viral replication; passive immunoprophylaxis; mutant.

XX Hepatitis C virus.

OS Synthetic.

XX Key Location/Qualifiers
 FH 1..20
 FT Peptide /label= Signal_peptide

FT /note= "Derived from endoplasmic reticulum E1
 signal sequence (364-383 amino acids)"

FT Protein 21..271

FT /label= Mature_C_terminal_truncated_E2_protein

FT Misc-difference 225

FT /note= "Wild type Leu substituted with His"

XX WO200121807-A1.

XX 29-MAR-2001.

XX 22-SEP-2000; 2000WO-US25987.

XX 23-SEP-1999; 99US-0155823.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Fornis X, Bukh J, Emerson SU, Purcell RH;

XX WPI; 2001-266076/27.

XX Novel nucleic acid molecules that encode hepatitis C virus envelope 2
 PT protein lacking all or part of hypervariable region 1 of envelope
 PT protein, useful as vaccine components for treating or preventing HCV
 PT infections -

XX Example; Page -: 80pp; English.

XX The present sequence is hepatitis C virus (HCV) E2-634 HVR1 construct
 CC containing the endoplasmic reticulum signal sequence of envelope protein

CC E1 and carboxy-terminal truncated E2 protein lacking the hypervariable
 CC region one (HVR1) with a replacement of leucine for histidine. The HCV
 CC E2 protein lacking HVR1 DNA is useful for producing infectious HCV and
 CC chimeric HCV viruses which are useful for identifying cell lines
 CC capable of supporting the replication of viruses. The infectious HCV and
 CC HVR1-chimeric HCV are used in the production of attenuated or inactivated
 CC vaccines which are useful for treating or preventing HCV in a mammal by
 CC immunisation. The host cells expressing the H77C(HVR1) DNA is useful as
 CC an immunogen to stimulate a protective immune response to HCV. The
 CC immunogens are useful for producing protective antibodies to HCV. The
 CC antibodies produced are used in passive immunoprophylaxis for treatment
 CC of diseases caused by HCV in animals, especially humans. The H77C(HVR1)
 CC DNA is also useful in gene therapy.
 CC Note: The present sequence is not shown in the specification but is
 CC derived from Hepatitis C virus envelope 2 protein lacking hypervariable
 CC region 1 (AAE00442) referred as SEQ ID NO: 2 and shown in figure 1.

XX
 SQ Sequence 271 AA;
 Query Match 100.0%; Score 20; DB 22; Length 271;
 Best Local Similarity 100.0%; Pred. No. 3.6e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCTPSPVVVGT 20
 DB 109 AKSVCGPVYCTPSPVVVGT 128
 |||||

RESULT 13
 AAR33997
 ID AAR33997 standard; Protein; 278 AA.
 XX
 AC AAR33997;
 XX
 DT 25-MAR-2003 (updated)
 DT 26-JUL-1993 (first entry)
 XX
 DE Th E2/NS1 protein.
 XX
 KW Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
 KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
 KW domain; immunological; cross-reactive; envelope protein; vaccine;
 KW gp53 (BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.
 XX
 OS Synthetic.
 XX
 PN WO9306126-A1.
 XX
 XX 01-APR-1993.
 XX
 XX 11-SEP-1992; 92WO-US07683.
 XX
 XX 13-SEP-1991; 91US-0759575.
 XX
 XX (CHIR) CHIRON CORP.
 XX
 XX Houghton M, Weiner AJ;
 XX
 XX WPI; 1993-117468/14.
 XX

PT Immuno-reactive hepatitis C virus polypeptide compens. - contg.
 PT at least 2 sequences from the first variable domain of distinct
 PT HCV isolates

PS Disclosure; Fig 3; 106pp; English.

XX The sequences given in AAR33992-002 represent a portion of the E2/NS1
 CC protein encoded by group I and group II HCV isolates, from amino acid
 CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
 CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
 CC 30 amino acids which shows large variation between nearly all isolates.
 CC This is an important immunoreactive domain. This putative envelope
 CC glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera

CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
 CC flaviviruses, both of which confer protective immunity in hosts
 CC vaccinated with these polypeptides. It has been discovered that a
 CC number of important HCV epitopes vary among viral isolates and that
 CC these epitopes can be mapped to specific domains. This meant that
 CC immunologically cross-reactive polypeptides which focus on variable
 CC rather than constant domains can be produced. See also AAQ39134-48
 CC and AAR33982-91.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 278 AA;

Query Match 100.0%; Score 20; DB 14; Length 278;
 Best Local Similarity 100.0%; Pred. No. 3.7e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCTPSPVVVGT 20
 DB 130 AKSVCGPVYCTPSPVVVGT 149
 |||||

RESULT 14
 AAW0929
 ID AAW0929 standard; Protein; 305 AA.
 XX
 AC AAW0929;
 XX

DT 04-NOV-1997 (first entry)

XX Recombinant HCV E2 antigen.

DE HCV; E2; antigen; non-secretor gene; protein secretion; vaccine;
 KW plasmid 577.

XX Chimeric hepatitis C virus;
 OS Chimeric synthetic.

XX Key Location/Qualifiers
 FT Peptide 1..19
 FT /label= Sig_peptide

FT Cleavage-site 20
 FT /note= "mammalian secretion signal
 FT peptidase cleavage site"

FT Protein 20..305
 FT /label= Mat_protein

FT Peptide 25..28
 FT /note= "N-terminal sequence of prourokinase"

XX WO9641179-A1.

XX 19-DEC-1996.

XX 05-JUN-1996; 96WO-US09345.

XX 07-JUN-1995; 95US-0478073.

XX (ABBO) ABBOTT LAB.

XX Lesniewski RR, Okasinski GF, Schaefer VG, Suhar TS;

XX WPI; 1997-108653/10.

XX N-PSDB, AAT13899.

PT New expression system for proteins, partic. HCV antigens - for use
 PT in assays for screening and prognostic applications and for use in
 PT vaccines

XX Example 1; Page 29-30; 40pp; English.

XX This sequence comprises the conceptual translation product of
 CC a hepatitis C virus E2 antigen expression cassette (AAT13899). The
 CC encoded E2 antigen is truncated at amino acid residue 644 of HCV,
 CC and contains an N-terminal sequence (SNEI) from human prourokinase

CC intended to promote signal protease processing, efficient secretion
 CC and final product stability. Glycosylated E2 can be expressed in
 CC mammalian host cells utilizing claimed plasmid 577. Fusion
 CC proteins produced by plasmid 577 are used in claimed assays and
 CC test kits for detecting anti-analyte antibody as well as in claimed
 CC vaccines for treatment of infection. The plasmid can be used to
 CC produce high levels of proteins that would not normally be
 CC expressed in mammalian cells due to the non-secretory nature of the
 CC gene. The system is esp. used for expressing HCV proteins, allowing
 CC proper processing, glycosylation and conformation of the viral
 CC protein.
 SQ Sequence 305 AA;

Query Match 100.0%; Score 20; DB 18; Length 305;
 Best Local Similarity 100.0%; Pred. No. 3.9e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFPTSPVVVGT 20
 |||||
 Db 140 AKSVCGPVYCFPTSPVVVGT 159

RESULT 15

AAR40118
 ID AAR40118 standard; Protein; 333 AA.

AC AAR40118;

DT 25-MAR-2003 (updated)
 DT 27-JAN-1994 (first entry)

DE HGH-HCV-E2 fusion protein expressed by pHCV-170.

KW Hepatitis C Virus; Non-A, non-B hepatitis Virus; HCV; NANBHV;
 KW human growth hormone; HGH; secretion signal; fusion protein;
 KW vaccine.

OS Chimeric Hepatitis C Virus.
 OS Chimeric Homo sapiens.

FT Key Location/Qualifiers
 FT Region 1..32
 FT /note= "HGH secretion signal"
 FT Region 33..333
 FT /note= "HCV-E2 (amino acids 384-684)"

XX WO9315193-A1.

PN 05-AUG-1993.

PD 29-JAN-1993; 93WO-US00907.

XX 31-JAN-1992; 92US-0830024.

XX (ABBO) ABBOTT LAB.

XX Bode SL, Casey JM, Desai SM, Devare SG, Frail DE;
 PI Yamaguchi J, Zeck BJ;

XX WPI; 1993-258673/32.
 DR N-PSDB; AAQ47195.

XX New plasmid pHCV-162 is a mammalian expression systems for HCV
 PT proteins - useful for diagnosing HCV infection and as vaccines
 PT for preventing HCV infection

XX Claim 10; Page 81-82; 100pp; English.

XX A sequence coding for the HCV E2 protein (amino acids 384-684 from
 CC HCV-infected chimpanzee CO isolate, see AAR40119) was generated using
 CC PCR. An EcoRI site was used to attach a synthetic oligonucleotide
 CC encoding the Human Growth Hormone secretion signal at the 5'-end of

CC the HCV sequence. The resultant plasmid was designated pHCV-170.
 CC AAR40118 is the sequence of the HGH-HCV-E2 fusion protein expressed
 CC by vector pHCV-170. See AAQ47192-Q47196.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX SQ Sequence 333 AA;

Query Match 100.0%; Score 20; DB 14; Length 333;
 Best Local Similarity 100.0%; Pred. No. 4.2e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFPTSPVVVGT 20
 |||||
 Db 148 AKSVCGPVYCFPTSPVVVGT 167

Search completed: November 21, 2003, 20:58:04
 Job time : 32.15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.6 Seconds
(without alignments)
79.832 Million cell updates/sec

Title: US-09-973-025-80
Perfect score: 20
Sequence: 1 AKSVCGPVYCFTPSPVVVGT 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-08-612-973-80
2	20	100.0	20	3	US-08-927-597-80
3	20	100.0	20	4	US-08-635-886C-36
4	20	100.0	278	1	US-08-440-103-15
5	20	100.0	278	1	US-08-440-542-15
6	20	100.0	278	1	US-08-231-368-15
7	20	100.0	278	1	US-08-440-210-15
8	20	100.0	278	4	US-09-046-604-15
9	20	100.0	305	3	US-08-478-073-2
10	20	100.0	333	1	US-08-453-552-12
11	20	100.0	333	2	US-08-710-637-12
12	20	100.0	333	5	PCT-US93-00907-12
13	20	100.0	337	1	US-08-188-281B-7
14	20	100.0	337	5	PCT-US94-07280-7
15	20	100.0	337	5	PCT-US95-01087-7
16	20	100.0	367	1	US-08-188-281B-9
17	20	100.0	367	1	US-08-453-552-6
18	20	100.0	367	2	US-08-710-637-6
19	20	100.0	367	5	PCT-US93-00907-6
20	20	100.0	367	5	PCT-US94-07280-9
21	20	100.0	367	5	PCT-US95-01087-9
22	20	100.0	377	1	US-08-188-281B-17
23	20	100.0	377	5	PCT-US94-07280-17
24	20	100.0	377	5	PCT-US95-01087-17
25	20	100.0	397	1	US-08-188-281B-11
26	20	100.0	397	5	PCT-US94-07280-11
27	20	100.0	397	5	PCT-US95-01087-11

28	20	100.0	399	1	US-08-453-552-10	Sequence 10, Appl
29	20	100.0	399	2	US-08-710-637-10	Sequence 10, Appl
30	20	100.0	399	5	PCT-US93-00907-10	Sequence 10, Appl
31	20	100.0	402	1	US-08-460-806-13	Sequence 13, Appl
32	20	100.0	402	1	US-08-460-806-15	Sequence 15, Appl
33	20	100.0	402	1	US-08-460-806-17	Sequence 17, Appl
34	20	100.0	402	1	US-08-325-630-13	Sequence 13, Appl
35	20	100.0	402	1	US-08-325-630-15	Sequence 15, Appl
36	20	100.0	402	1	US-08-325-630-17	Sequence 17, Appl
37	20	100.0	403	2	US-08-483-695-39	Sequence 39, Appl
38	20	100.0	403	2	US-07-985-285-39	Sequence 39, Appl
39	20	100.0	403	2	US-08-487-231-39	Sequence 39, Appl
40	20	100.0	403	3	US-09-201-912-39	Sequence 39, Appl
41	20	100.0	409	1	US-08-440-103-21	Sequence 21, Appl
42	20	100.0	409	1	US-08-440-103-24	Sequence 24, Appl
43	20	100.0	409	1	US-08-440-542-21	Sequence 21, Appl
44	20	100.0	409	1	US-08-440-542-24	Sequence 24, Appl
45	20	100.0	409	1	US-08-231-368-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-08-612-973-80
; Sequence 80, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-612-973-80

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AKSVCGPVYCFTPSPVVVGT 20
|||||

Db 1 AKSVCGPVYCFTSPVVGVT 20

RESULT 2

US-08-927-597-80
; Sequence 80, Application US/08927597
; Patent No. 6245503

GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

US-08-927-597-80

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFTSPVVGVT 20

Db 1 AKSVCGPVYCFTSPVVGVT 20

RESULT 3

US-08-635-886C-36
; Sequence 36, Application US/08635886C
; Patent No. 6555114

GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEUX, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555

; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 36
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus

US-08-635-886C-36

Query Match 100.0%; Score 20; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFTSPVVGVT 20

Db 1 AKSVCGPVYCFTSPVVGVT 20

RESULT 4

US-08-440-103-15
; Sequence 15, Application US/08440103
; Patent No. 5670152

GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,103
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575

ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-440-103-15

Query Match 100.0%; Score 20; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.9e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFTSPVVGVT 20

Db 130 AKSVCGPVYCFTSPVVGVT 149

```
RESULT 5
US-08-440-542-15
; Sequence 15, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,542
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-542-15
Query Match 100.0%; Score 20; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.9e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20
Db 130 AKSVCGPVYCFPTSPVVVGT 149

RESULT 6
US-08-231-368-15
; Sequence 15, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,210
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-542-15
Query Match 100.0%; Score 20; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.9e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20
Db 130 AKSVCGPVYCFPTSPVVVGT 149

RESULT 7
US-08-440-210-15
; Sequence 15, Application US/08440210
; Patent No. 5766845
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,210
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-231-368-15
Query Match 100.0%; Score 20; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.9e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20
Db 130 AKSVCGPVYCFPTSPVVVGT 149

RESULT 7
US-08-440-210-15
; Sequence 15, Application US/08440210
; Patent No. 5766845
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,210
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
```

TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 278 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-210-15

Query Match 100.0%; Score 20; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.9e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFTSPVVGVT 20
|||||
Db 130 AKSVCGPVYCFTSPVVGVT 149

RESULT 8

US-09-046-604-15
Sequence 15, Application US/09046604
Patent No. 6303292
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/046,604
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
FILING DATE:
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 278 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-046-604-15

Query Match 100.0%; Score 20; DB 4; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.9e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFTSPVVGVT 20
|||||
Db 130 AKSVCGPVYCFTSPVVGVT 149

RESULT 9

US-08-478-073-2
Sequence 2, Application US/08478073
Patent No. 6020122
GENERAL INFORMATION:
APPLICANT: Okasinski, Gregory F.
APPLICANT: Schaefer, Verlyn G.
APPLICANT: Suhar, Thomas S.
APPLICANT: Lesniewski, Richard R.
TITLE OF INVENTION: EXPRESSION SYSTEM FOR NON-SECRETOR GENES
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES
STREET: ONE HUNDRED ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,073
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Porembski, Priscilla E.
REGISTRATION NUMBER: 33207
REFERENCE/DOCKET NUMBER: 5763.US.01
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-073-2

Query Match 100.0%; Score 20; DB 3; Length 305;
Best Local Similarity 100.0%; Pred. No. 2.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFTSPVVGVT 20
|||||
Db 140 AKSVCGPVYCFTSPVVGVT 159

RESULT 10

US-08-453-552-12
Sequence 12, Application US/08453552
Patent No. 5667992
GENERAL INFORMATION:
APPLICANT: CASEY, JAMES M.
APPLICANT: BODE, SUZANNE L.
APPLICANT: ZECK, BILLY J.
APPLICANT: YAMAGUCHI, JULIE
APPLICANT: FRAIL, DONALD E.
APPLICANT: DESAI, SURESH M.
APPLICANT: DEVARE, SUSHIL G.
TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/453.552
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FOREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5131.US.D1
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 333 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-453-552-12

Query Match 100.0%; Score 20; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCTPSPVVVGT 20
Db 148 AKSVCGPVYCTPSPVVVGT 167

RESULT 11

US-08-710-637-12
Sequence 12, Application US/08710637
Patent No. 5854001
GENERAL INFORMATION:
APPLICANT: CASEY, JAMES M.
APPLICANT: BODE, SUZANNE L.
APPLICANT: ZECK, BILLY J.
APPLICANT: YAMAGUCHI, JULIE
APPLICANT: FRAIL, DONALD E.
APPLICANT: DESAI, SURESH M.
APPLICANT: DEVARE, SUSHIL G.
TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: US/08/710,637
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/144,099
FILING DATE:
APPLICATION NUMBER: US 07/830,024
FILING DATE: 01-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: FOREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5131.US.01
TELECOMMUNICATION INFORMATION:

TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 333 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-710-637-12

Query Match 100.0%; Score 20; DB 2; Length 333;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCTPSPVVVGT 20
Db 148 AKSVCGPVYCTPSPVVVGT 167

RESULT 12

PCT-US93-00907-12
Sequence 12, Application PC/TUS9300907
GENERAL INFORMATION:
APPLICANT: CASEY, JAMES M.
APPLICANT: BODE, SUZANNE L.
APPLICANT: ZECK, BILLY J.
APPLICANT: YAMAGUCHI, JULIE
APPLICANT: FRAIL, DONALD E.
APPLICANT: DESAI, SURESH M.
APPLICANT: DEVARE, SUSHIL G.
TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/00907
FILING DATE: 19930129
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: FOREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5131.PC.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 333 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-00907-12

Query Match 100.0%; Score 20; DB 5; Length 333;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCTPSPVVVGT 20
Db 148 AKSVCGPVYCTPSPVVVGT 167

RESULT 13
PCT-US94-07280-7
; Sequence 7, Application US/08188281B
; Patent No. 5610009
; GENERAL INFORMATION:
; APPLICANT: WATANABE, SHINICHI
; APPLICANT: YAMAGUCHI, JULIE
; APPLICANT: DESAI, SURESH M.
; APPLICANT: DEVARE, SUSHIL G.
; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
; TITLE OF INVENTION: ENVELOPE GENES
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,281B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5521.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-188-281B-7
Query Match 100.0%; Score 20; DB 1; Length 337;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AKSVCGPVYCFPSPVVGT 20
Db 182 AKSVCGPVYCFPSPVVGT 201

RESULT 14
PCT-US94-07280-7
; Sequence 7, Application PC/TUS9407280
; GENERAL INFORMATION:
; APPLICANT: WATANABE, SHINICHI
; APPLICANT: YAMAGUCHI, JULIE
; APPLICANT: DESAI, SURESH M.
; APPLICANT: DEVARE, SUSHIL G.
; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
; TITLE OF INVENTION: ENVELOPE GENES
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07280
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5521.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-07280-7
Query Match 100.0%; Score 20; DB 5; Length 337;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AKSVCGPVYCFPSPVVGT 20
Db 182 AKSVCGPVYCFPSPVVGT 201

RESULT 15
PCT-US99-01087-7
; Sequence 7, Application PC/TUS9901087
; GENERAL INFORMATION:
; APPLICANT: WATANABE, SHINICHI
; APPLICANT: YAMAGUCHI, JULIE
; APPLICANT: DESAI, SURESH M.
; APPLICANT: DEVARE, SUSHIL G.
; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
; TITLE OF INVENTION: ENVELOPE GENES
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01087
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5521.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

PCT-US95-01087-7

Query Match 100.0%; Score 20; DB 5; Length 337;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFPSFVVGT 20
|||
Db 182 AKSVCGPVYCFPSFVVGT 201
|||

Search completed: November 21, 2003, 21:15:16
Job time : 11.6 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time 19.5 Seconds
(without alignments)
187,240 Million cell updates/sec

Title: US-09-973-025-80

Perfect score: 20

Sequence: 1 AKSVCPVYCFTPSPVVGVT 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	10	US-09-973-025-80
2	20	100.0	20	11	US-09-899-303-80
3	20	100.0	20	11	US-09-995-808-80
4	20	100.0	20	11	US-09-995-860-80
5	20	100.0	20	12	US-09-995-791-80
6	20	100.0	43	10	US-09-921-397-22
7	20	100.0	96	10	US-09-921-397-82
8	20	100.0	176	10	US-09-921-397-81
9	20	100.0	250	10	US-09-952-572-80
10	20	100.0	254	10	US-09-407-430-3
11	20	100.0	350	10	US-09-929-955-4
12	20	100.0	350	14	US-10-104-966-4
13	20	100.0	363	10	US-09-407-430-2
14	20	100.0	363	12	US-10-128-587A-97
15	20	100.0	363	15	US-10-128-590-97

16	20	100.0	637	12	US-10-187-257-4	Sequence 4, Appli
17	20	100.0	637	12	US-10-265-083-2	Sequence 2, Appli
18	20	100.0	2894	10	US-09-941-611-23	Sequence 23, Appli
19	20	100.0	2894	15	US-10-044-995-23	Sequence 23, Appli
20	20	100.0	3011	9	US-09-742-659-4	Sequence 4, Appli
21	20	100.0	3011	9	US-09-916-359-2	Sequence 2, Appli
22	20	100.0	3011	10	US-09-238-076-20	Sequence 20, Appli
23	20	100.0	3011	10	US-09-952-572-9	Sequence 9, Appli
24	20	100.0	3011	10	US-09-929-955-1	Sequence 1, Appli
25	20	100.0	3011	10	US-09-747-419-20	Sequence 20, Appli
26	20	100.0	3011	11	US-09-891-894-3	Sequence 3, Appli
27	20	100.0	3011	11	US-09-995-937-20	Sequence 20, Appli
28	20	100.0	3011	11	US-09-917-563-20	Sequence 20, Appli
29	20	100.0	3011	12	US-10-184-150-3	Sequence 3, Appli
30	20	100.0	3011	14	US-10-104-966-1	Sequence 1, Appli
31	20	100.0	3011	15	US-10-259-275-20	Sequence 20, Appli
32	20	100.0	3011	16	US-10-232-643-6	Sequence 6, Appli
33	20	100.0	3012	10	US-09-238-076-2	Sequence 2, Appli
34	20	100.0	3012	11	US-09-995-937-2	Sequence 2, Appli
35	20	100.0	3012	11	US-09-917-563-2	Sequence 2, Appli
36	17	85.0	26	10	US-09-921-397-7	Sequence 7, Appli
37	17	85.0	28	10	US-09-921-397-19	Sequence 19, Appli
38	17	85.0	31	16	US-10-318-200-13	Sequence 13, Appli
39	17	85.0	290	12	US-10-128-587A-3	Sequence 3, Appli
40	17	85.0	290	15	US-10-128-590-3	Sequence 3, Appli
41	17	85.0	301	12	US-10-128-587A-5	Sequence 5, Appli
42	17	85.0	301	15	US-10-128-590-5	Sequence 5, Appli
43	17	85.0	314	10	US-09-973-025-42	Sequence 42, Appli
44	17	85.0	314	11	US-09-899-303-42	Sequence 42, Appli
45	17	85.0	314	11	US-09-995-808-42	Sequence 42, Appli

ALIGNMENTS

RESULT 1

US-09-973-025-80
; Sequence 80, Application US/09973025
; Publication No. US20020182706A1

; GENERAL INFORMATION:

; APPLICANT: MAERTENS, GEERT

; BOSMAN, FONS

; DE MARTYNOFF, GUY

; BUYSSE, MARIE-ANGE

; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

; NUMBER OF SEQUENCES: 111

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: NIXON & VANDERHUYE P.C.

; 0 STREET: 1100 NORTH GLEBE ROAD

; CITY: ARLINGTON

; STATE: VIRGINIA

; COUNTRY: U.S.A.

; ZIP: 22201-4714

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION NUMBER: US/09/973,025

; FILING DATE: 10-Oct-2001

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/612,973

; FILING DATE: 11-MAR-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: BYRNE, THOMAS E.

; REGISTRATION NUMBER: 32,205

; REFERENCE/DOCKET NUMBER: 1487-10

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 816-4000

; TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 80:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 80:
US-09-973-025-80

Query Match 100.0%; Score 20; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20
| | | | | | | | | | | | | | | | | | | | | |
DB 1 AKSVCGPVYCFPTSPVVVGT 20

RESULT 2
US-09-899-303-80
; Sequence 80, Application US/09899303
; Publication No. US20030036110A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 80:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 80:
US-09-899-303-80

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20
| | | | | | | | | | | | | | | | | | | | | |

DB 1 AKSVCGPVYCFPTSPVVVGT 20

RESULT 3
US-09-995-808-80
; Sequence 80, Application US/09995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 80
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-80

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20
| | | | | | | | | | | | | | | | | | | | | |
DB 1 AKSVCGPVYCFPTSPVVVGT 20

RESULT 4
US-09-995-860-80
; Sequence 80, Application US/09995860
; Publication No. US20030118603A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 80
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-80

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20
| | | | | | | | | | | | | | | | | | | | | |
DB 1 AKSVCGPVYCFPTSPVVVGT 20

RESULT 5
US-09-995-791-80
; Sequence 80, Application US/09995791
; Publication No. US20030147918A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 80

; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-80

Query Match 100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFTSPVVGVT 20
Db 1 AKSVCGPVYCFTSPVVGVT 20
|||||

RESULT 6

US-09-921-397-22
; Sequence 22, Application US/09921397
; Patent No. US20020151484A1

; GENERAL INFORMATION:

; APPLICANT: HYBRIGENICS

; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof

; FILE REFERENCE: B4809A - JAZ

; CURRENT APPLICATION NUMBER: US/09/921,397

; PRIOR FILING DATE: 2001-08-02

; PRIOR APPLICATION NUMBER: EP 00402225.7

; PRIOR FILING DATE: 2000-08-03

; NUMBER OF SEQ ID NOS: 156

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 22

; LENGTH: 43

; TYPE: PRT

; ORGANISM: Hepatitis C virus

US-09-921-397-22

Query Match 100.0%; Score 20; DB 10; Length 43;
Best Local Similarity 100.0%; Pred. No. 2.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFTSPVVGVT 20
Db 5 AKSVCGPVYCFTSPVVGVT 24
|||||

RESULT 7

US-09-921-397-82
; Sequence 82, Application US/09921397
; Patent No. US20020151484A1

; GENERAL INFORMATION:

; APPLICANT: HYBRIGENICS

; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof

; FILE REFERENCE: B4809A - JAZ

; CURRENT APPLICATION NUMBER: US/09/921,397

; PRIOR FILING DATE: 2001-08-02

; PRIOR APPLICATION NUMBER: EP 00402225.7

; PRIOR FILING DATE: 2000-08-03

; NUMBER OF SEQ ID NOS: 156

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 82

; LENGTH: 96

; TYPE: PRT

; ORGANISM: Hepatitis C virus

US-09-921-397-82

Query Match 100.0%; Score 20; DB 10; Length 96;
Best Local Similarity 100.0%; Pred. No. 4.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFTSPVVGVT 20
|||||

Db 13 AKSVCGPVYCFTSPVVGVT 32

RESULT 8

US-09-921-397-81
; Sequence 81, Application US/09921397
; Patent No. US20020151484A1

; GENERAL INFORMATION:

; APPLICANT: HYBRIGENICS

; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof

; FILE REFERENCE: B4809A - JAZ

; CURRENT APPLICATION NUMBER: US/09/921,397

; PRIOR FILING DATE: 2001-08-02

; PRIOR APPLICATION NUMBER: EP 00402225.7

; PRIOR FILING DATE: 2000-08-03

; NUMBER OF SEQ ID NOS: 156

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 81

; LENGTH: 176

; TYPE: PRT

; ORGANISM: Hepatitis C virus

US-09-921-397-81

Query Match 100.0%; Score 20; DB 10; Length 176;

Best Local Similarity 100.0%; Pred. No. 8e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFTSPVVGVT 20

Db 31 AKSVCGPVYCFTSPVVGVT 50

RESULT 9

US-09-952-572-8
; Sequence 8, Application US/09952572
; Patent No. US20020119495A1

; GENERAL INFORMATION:

; APPLICANT: HAWAII BIOTECHNOLOGY GROUP, Inc.

; APPLICANT: NAKANO, Eileen

; APPLICANT: CLEMENTS, David

; APPLICANT: HUMPHREYS, Tom

; TITLE OF INVENTION: IMMUNOGENIC COMPOSITION OF HEPATITIS C

; FILE REFERENCE: HAWBIO1100

; CURRENT APPLICATION NUMBER: US/09/952,572

; CURRENT FILING DATE: 2001-09-13

; PRIOR APPLICATION NUMBER: US 60/230,927

; PRIOR FILING DATE: 2000-09-13

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 8

; LENGTH: 250

; TYPE: PRT

; ORGANISM: Hepatitis C Virus

US-09-952-572-8

Query Match 100.0%; Score 20; DB 10; Length 250;

Best Local Similarity 100.0%; Pred. No. 1.1e-12;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFTSPVVGVT 20

Db 88 AKSVCGPVYCFTSPVVGVT 107

RESULT 10

US-09-407-430-3

; Sequence 3, Application US/09407430

; Patent No. US20020160936A1

; GENERAL INFORMATION:

; APPLICANT: Worman, Howard J.

; APPLICANT: Mamiya, Naoto

; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-407-430-3

Query Match 100.0%; Score 20; DB 10; Length 254;
Best Local Similarity 100.0%; Pred. No. 1.1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFPTSPVVVGT 20
| | | | | | | | | | | | | | | | | | | | | |
Db 94 AKSVCGPVYCFPTSPVVVGT 113

RESULT 11
US-09-929-955-4
; Sequence 4, Application US/09929955
; Patent No. US20020136740A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; FILE REFERENCE: TRIPEP.23AUS2
; CURRENT APPLICATION NUMBER: US/09/929,955
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/225,767
; PRIOR FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-09-929-955-4

Query Match 100.0%; Score 20; DB 10; Length 350;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFPTSPVVVGT 20
| | | | | | | | | | | | | | | | | | | | | |
Db 120 AKSVCGPVYCFPTSPVVVGT 139

RESULT 12
US-10-104-966-4
; Sequence 4, Application US/10104966
; Publication No. US20020155124A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; FILE REFERENCE: TRIPEP.23AUSC1
; CURRENT APPLICATION NUMBER: US/10/104,966
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03

; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-10-104-966-4

Query Match 100.0%; Score 20; DB 14; Length 350;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFPTSPVVVGT 20
| | | | | | | | | | | | | | | | | | | | | |
Db 120 AKSVCGPVYCFPTSPVVVGT 139

RESULT 13
US-09-407-430-2
; Sequence 2, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Norman, Howard J.
; APPLICANT: Mamuya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-09-407-430-2

Query Match 100.0%; Score 20; DB 10; Length 363;
Best Local Similarity 100.0%; Pred. No. 1.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFPTSPVVVGT 20
| | | | | | | | | | | | | | | | | | | | | |
Db 116 AKSVCGPVYCFPTSPVVVGT 135

RESULT 14
US-10-128-587A-97
; Sequence 97, Application US/10128587A
; Publication No. US20030152940A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Constructs and methods for expression of recombinant HCV envelope
; FILE REFERENCE: 134 PCT
; CURRENT APPLICATION NUMBER: US/10/128,587A
; CURRENT FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: hepatitis C virus
US-10-128-587A-97

Query Match 100.0%; Score 20; DB 12; Length 363;
Best Local Similarity 100.0%; Pred. No. 1.5e-12;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20
|||||
Db 116 AKSVCGPVYCFPTSPVVVGT 135

RESULT 15

US-10-128-590-97
; Sequence 97, Application US/10128590
; Publication No. US20030108561A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Core-glycosylated HCV envelope proteins,
; FILE REFERENCE: 135 PCT
; CURRENT APPLICATION NUMBER: US/10/128,590
; CURRENT FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-128-590-97

Query Match 100.0%; Score 20; DB 15; Length 363;
Best Local Similarity 100.0%; Pred. No. 1.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20
|||||
Db 116 AKSVCGPVYCFPTSPVVVGT 135

Search completed: November 21, 2003, 22:19:38
Job time : 19.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 163.1 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-80
Perfect score: 20
Sequence: 1 AKSCVGPVCFPTSPVVGVT 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0
Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA Main.*
1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US097A_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US097B_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US099A_COMB.pep.*
25: /cgn2_6/ptodata/1/paa/US099B_COMB.pep.*
26: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
27: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
28: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
29: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
30: /cgn2_6/ptodata/1/paa/US104_COMB.pep.*
31: /cgn2_6/ptodata/1/paa/US106_COMB.pep.*
32: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	20	100.0	20 13	US-08-974-685-36
				Sequence 36, Appl

2	20	100.0	20 13	US-08-974-690-36	Sequence 36, Appl
3	20	100.0	20 13	US-08-974-690A-36	Sequence 36, Appl
4	20	100.0	20 13	US-08-974-690B-36	Sequence 36, Appl
5	20	100.0	20 13	US-08-974-690C-36	Sequence 36, Appl
6	20	100.0	20 23	US-09-899-303-80	Sequence 80, Appl
7	20	100.0	20 23	US-09-899-303A-80	Sequence 80, Appl
8	20	100.0	20 25	US-09-973-025-80	Sequence 80, Appl
9	20	100.0	20 25	US-09-973-025-80	Sequence 80, Appl
10	20	100.0	20 25	US-09-995-791-80	Sequence 80, Appl
11	20	100.0	20 25	US-09-995-808-80	Sequence 80, Appl
12	20	100.0	20 25	US-09-995-860-80	Sequence 80, Appl
13	20	100.0	20 26	US-10-020-510-80	Sequence 80, Appl
14	20	100.0	20 29	US-10-321-798-80	Sequence 80, Appl
15	20	100.0	43 24	US-09-921-397-22	Sequence 22, Appl
16	20	100.0	96 24	US-09-921-397-82	Sequence 82, Appl
17	20	100.0	122 8	US-08-436-966-5	Sequence 5, Appl
18	20	100.0	176 24	US-09-921-397-81	Sequence 81, Appl
19	20	100.0	250 1	PCT-US01-28767-8	Sequence 8, Appl
20	20	100.0	250 25	US-09-952-572-8	Sequence 8, Appl
21	20	100.0	254 1	PCT-US00-26395-3	Sequence 3, Appl
22	20	100.0	254 18	US-09-407-430-3	Sequence 3, Appl
23	20	100.0	278 8	US-08-471-498-15	Sequence 15, Appl
24	20	100.0	333 8	US-08-417-478-12	Sequence 12, Appl
25	20	100.0	333 8	US-08-453-613-12	Sequence 3, Appl
26	20	100.0	333 30	US-10-445-724-3	Sequence 3, Appl
27	20	100.0	333 32	US-09-409-909-4	Sequence 4, Appl
28	20	100.0	350 21	US-09-705-547-4	Sequence 4, Appl
29	20	100.0	350 24	US-09-929-955-4	Sequence 4, Appl
30	20	100.0	350 27	US-10-104-966-4	Sequence 4, Appl
31	20	100.0	350 32	US-09-229-175-4	Sequence 4, Appl
32	20	100.0	363 1	PCT-US00-26395-2	Sequence 2, Appl
33	20	100.0	363 18	US-09-407-430-2	Sequence 2, Appl
34	20	100.0	363 27	US-10-128-587A-97	Sequence 97, Appl
35	20	100.0	367 8	US-08-417-478-6	Sequence 6, Appl
36	20	100.0	367 8	US-08-453-613-6	Sequence 6, Appl
37	20	100.0	399 8	US-08-417-478-10	Sequence 10, Appl
38	20	100.0	399 8	US-08-453-613-10	Sequence 10, Appl
39	20	100.0	409 8	US-08-471-498-21	Sequence 21, Appl
40	20	100.0	409 8	US-08-471-498-24	Sequence 24, Appl
41	20	100.0	416 8	US-08-436-966-9	Sequence 9, Appl
42	20	100.0	421 29	US-10-365-620-70	Sequence 70, Appl
43	20	100.0	454 8	US-08-403-590B-73	Sequence 73, Appl
44	20	100.0	454 8	US-08-444-112-73	Sequence 73, Appl
45	20	100.0	463 3	US-07-748-561-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-974-685-36
; Sequence 36, Application US/08974685
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; MAERTENS, ROBERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
; THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 181
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,685
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 36:

US-08-974-685-36

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFTSPSPVVVGT 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 AKSVCGPVYCFTSPSPVVVGT 20

RESULT 2

US-08-974-690-36

; Sequence 36, Application US/08974690
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; APPLICANT: DELEYS, ROBERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; TITLE OF INVENTION: HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
; TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 166
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P. C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE: 25-APR-1996
; APPLICATION NUMBER: US/08/974,690

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/635,886

FILING DATE: 25-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-11

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 36:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-974-690-36

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFTSPSPVVVGT 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 AKSVCGPVYCFTSPSPVVVGT 20

RESULT 3

US-08-974-690A-36

; Sequence 36, Application US/08974690A
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; APPLICANT: DELEYS, ROBERT
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; TITLE OF INVENTION: HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
; TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P. C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE: 19-Nov-1997
; APPLICATION NUMBER: US/08/974,690A
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B. J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 2752-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 36:

US-08-974-690A-36

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFTSPSPVVVGT 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 AKSVCGPVYCFTSPSPVVVGT 20

RESULT 4

US-08-974-690B-36

; Sequence 36, Application US/08974690B
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; APPLICANT: DELEYS, ROBERT
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; TITLE OF INVENTION: HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
; TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 166
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P. C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE: 25-APR-1996
; APPLICATION NUMBER: US/08/974,690

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/635,886

FILING DATE: 25-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-11

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 36:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

MOLECULE TYPE: peptide

THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHUYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690B
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-08-974-690B-36

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFPSPPVVVGT 20
|||||

Db 1 AKSVCGPVYCFPSPPVVVGT 20
|||||

RESULT 5

US-08-974-690C-36
Sequence 36, Application US/08974690C
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELEYS, Robert
APPLICANT: MAERTENS, Geert
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
FILE REFERENCE: 2551-94
CURRENT APPLICATION NUMBER: US/08/974,690C
CURRENT FILING DATE: 1997-11-19
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28
PRIOR APPLICATION NUMBER: EP 93402718.6
PRIOR FILING DATE: 1993-11-04
NUMBER OF SEQ ID NOS: 286
SOFTWARE: PatentIn version 3.1
SEQ ID NO 36
LENGTH: 20
TYPE: PPT
ORGANISM: hepatitis C virus
US-08-974-690C-36

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFPSPPVVVGT 20

Db 1 AKSVCGPVYCFPSPPVVVGT 20
|||||

RESULT 6

US-09-899-303-80
Sequence 80, Application US/09899303
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHUYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 80:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 80:
US-09-899-303-80

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFPSPPVVVGT 20
|||||

Db 1 AKSVCGPVYCFPSPPVVVGT 20
|||||

RESULT 7

US-09-899-303A-80
Sequence 80, Application US/09899303A
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHUYE P.C.
US-09-899-303A-80

STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/999,303A
FILING DATE: 06-Jul-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 80:
SEQUENCE DESCRIPTION: SEQ ID NO: 80:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 80:
US-09-899-303A-80
Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AKSVCGPVYCFTSPVVVGT 20
Db 1 AKSVCGPVYCFTSPVVVGT 20
RESULT 8
US-09-973-025-80
; Sequence 80, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; RUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 80:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 80:
US-09-973-025-80
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AKSVCGPVYCFTSPVVVGT 20
Db 1 AKSVCGPVYCFTSPVVVGT 20
RESULT 9
US-09-995-791-80
; Sequence 80, Application US/09995791
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; therapeutic use.
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 80
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; US-09-995-791-80
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AKSVCGPVYCFTSPVVVGT 20
Db 1 AKSVCGPVYCFTSPVVVGT 20
RESULT 10
US-09-995-808-80
; Sequence 80, Application US/09995808
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 80
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; US-09-995-808-80
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.1e-13;

```
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AKSVCGPVYCFTSPSPVVVGT 20
    |||||
Db 1 AKSVCGPVYCFTSPSPVVVGT 20
    |||||

RESULT 11
US-09-995-860-80
; Sequence 80, Application US/09995860
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 80
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-80

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AKSVCGPVYCFTSPSPVVVGT 20
    |||||
Db 1 AKSVCGPVYCFTSPSPVVVGT 20
    |||||

RESULT 12
US-10-020-510-80
; Sequence 80, Application US/10020510
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-72
; CURRENT APPLICATION NUMBER: US/10/020,510
; CURRENT FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 80
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-020-510-80

Query Match 100.0%; Score 20; DB 26; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AKSVCGPVYCFTSPSPVVVGT 20
    |||||
Db 1 AKSVCGPVYCFTSPSPVVVGT 20
    |||||

RESULT 13
US-10-321-798-80
; Sequence 80, Application US/10321798
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-93
; CURRENT APPLICATION NUMBER: US/10/321,798
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: 60/418,358
; PRIOR FILING DATE: 2002-10-16
```

```
; PRIOR APPLICATION NUMBER: 10/020,510
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 80
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-321-798-80

Query Match 100.0%; Score 20; DB 29; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AKSVCGPVYCFTSPSPVVVGT 20
    |||||
Db 1 AKSVCGPVYCFTSPSPVVVGT 20
    |||||

RESULT 14
US-09-921-397-22
; Sequence 22, Application US/09921397
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 43
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-22

Query Match 100.0%; Score 20; DB 24; Length 43;
Best Local Similarity 100.0%; Pred. No. 1.8e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AKSVCGPVYCFTSPSPVVVGT 20
    |||||
Db 5 AKSVCGPVYCFTSPSPVVVGT 24
    |||||

RESULT 15
US-09-921-397-82
; Sequence 82, Application US/09921397
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 82
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-82

Query Match 100.0%; Score 20; DB 24; Length 96;
Best Local Similarity 100.0%; Pred. No. 3.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 AKSVCGPVYCFPTSPVVVGT 20
| | | | | | | | | | | | | | | | | | | | | |
Db 13 AKSVCGPVYCFPTSPVVVGT 32

Search completed: November 21, 2003, 22:09:52
Job time : 164.1 secs

0 1

0 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.55 Seconds
(without alignments)
117.536 Million cell updates/sec

Title: US-09-973-025-80

Perfect score: 20

Sequence: 1 AKSVCGPVYCFPTSPVVGVT 20

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- 1: Pending Patents_AA_New.*
- 2: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	20	100.0	20	6 US-10-651-165-36
2	20	100.0	30	6 US-10-296-734-474
3	20	100.0	333	1 PCT-US03-19834-3
4	20	100.0	333	6 US-10-655-562-4
5	20	100.0	637	1 PCT-US03-33610-4
6	20	100.0	2010	6 US-10-296-734-814
7	20	100.0	3011	1 PCT-US03-19834-2
8	20	100.0	3011	6 US-10-296-734-406
9	20	100.0	5985	6 US-10-296-734-810
10	17	85.0	31	6 US-10-685-435-22
11	17	85.0	347	6 US-10-664-391-9
12	17	85.0	539	6 US-10-664-391-11
13	17	85.0	1026	1 PCT-US03-20409-3
14	17	85.0	2280	1 PCT-US03-20322-211
15	17	85.0	3033	6 US-10-009-002-5
16	12	60.0	34	6 US-10-685-435-21
17	10	50.0	30	6 US-10-296-734-472
18	10	50.0	30	6 US-10-296-734-476
19	10	50.0	2011	6 US-10-296-734-812
20	8	40.0	20	6 US-10-651-165-37
21	8	40.0	20	6 US-10-651-165-37
22	6	30.0	168	6 US-10-389-647-431
23	6	30.0	219	6 US-10-425-114A-55710
24	6	30.0	297	7 US-60-490-890-976
25	6	30.0	302	6 US-10-425-114A-5114
26	6	30.0	320	1 PCT-US03-28227-3015

27	6	30.0	373	1	PCT-US03-28227-3016	Sequence 3016, Ap
28	6	30.0	606	6	US-10-425-114A-69305	Sequence 69305, A
29	5	25.0	20	1	PCT-US02-32947B-6	Sequence 6, Appli
30	5	25.0	21	6	US-10-475-117-285	Sequence 285, App
31	5	25.0	26	6	US-10-670-186-759	Sequence 759, App
32	5	25.0	27	5	US-09-538-038A-893	Sequence 893, App
33	5	25.0	55	6	US-10-670-186-137	Sequence 137, App
34	5	25.0	55	6	US-10-670-186-472	Sequence 472, App
35	5	25.0	69	6	US-10-425-114A-55965	Sequence 55965, A
36	5	25.0	83	6	US-10-425-114A-58169	Sequence 58169, A
37	5	25.0	107	5	US-09-979-932A-886	Sequence 886, App
38	5	25.0	107	5	US-09-979-932A-887	Sequence 887, App
39	5	25.0	107	5	US-09-979-932A-889	Sequence 889, App
40	5	25.0	107	5	US-09-979-932A-891	Sequence 891, App
41	5	25.0	109	1	PCT-US03-25002-1	Sequence 1, Appli
42	5	25.0	109	1	PCT-US03-25002-2	Sequence 2, Appli
43	5	25.0	112	6	US-10-679-063-3169	Sequence 3169, Ap
44	5	25.0	113	6	US-10-679-063-14130	Sequence 14130, A
45	5	25.0	117	6	US-10-679-063-3168	Sequence 3168, Ap

ALIGNMENTS

RESULT 1
US-10-651-165-36
; Sequence 36, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEMS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-36

Query Match 100.0%; Score 20; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVGVT 20
DB 1 AKSVCGPVYCFPTSPVVGVT 20

RESULT 2
US-10-296-734-474
; Sequence 474, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2

```
; SEQ ID NO 474
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 34
US-10-296-734-474

Query Match      100.0%; Score 20; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 5.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCTPSPVVVGT 20
Db 6 AKSVCGPVYCTPSPVVVGT 25

RESULT 3
PCT-US03-19834-3
; Sequence 3, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; TITLE OF INVENTION: METHODS FOR TREATING HYPERLIPIDEMIA
; FILE REFERENCE: IOWA:045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-3

Query Match      100.0%; Score 20; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 2.6e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCTPSPVVVGT 20
Db 117 AKSVCGPVYCTPSPVVVGT 136

RESULT 4
US-10-655-562-4
; Sequence 4, Application US/10655562
; GENERAL INFORMATION:
; APPLICANT: BISHOP-HURLEY, SHARON L.
; APPLICANT: SCHMIDT, FRANCIS J.
; APPLICANT: SMITH, ARNOLD L.
; TITLE OF INVENTION: PHAGE-DISPLAY PEPTIDES AS NOVEL ANTIMICROBIAL AGENTS
; TITLE OF INVENTION: AGAINST HAEMOPHILUS INFLUENZAE
; FILE REFERENCE: UVM0:022US
; CURRENT APPLICATION NUMBER: US/10/655,562
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: 60,409,909
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-655-562-4
```

```
Query Match      100.0%; Score 20; DB 6; Length 333;
Best Local Similarity 100.0%; Pred. No. 2.6e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCTPSPVVVGT 20
Db 117 AKSVCGPVYCTPSPVVVGT 136

RESULT 5
PCT-US03-33610-4
; Sequence 4, Application PC/TUS0333610
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: SELBY, Mark
; APPLICANT: PALIARD, Xavier
; TITLE OF INVENTION: ACTIVATION OF HCV-SPECIFIC T-CELLS
; FILE REFERENCE: 2300-1612.60
; CURRENT APPLICATION NUMBER: PCT/US03/33610
; CURRENT FILING DATE: 2003-10-24
; PRIOR APPLICATION NUMBER: 10/281,341
; PRIOR FILING DATE: 2002-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
PCT-US03-33610-4

Query Match      100.0%; Score 20; DB 1; Length 637;
Best Local Similarity 100.0%; Pred. No. 3.9e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCTPSPVVVGT 20
Db 327 AKSVCGPVYCTPSPVVVGT 346

RESULT 6
US-10-296-734-814
; Sequence 814, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 814
; LENGTH: 2010
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette B
US-10-296-734-814

Query Match      100.0%; Score 20; DB 6; Length 2010;
Best Local Similarity 100.0%; Pred. No. 8.4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCTPSPVVVGT 20
Db 1056 AKSVCGPVYCTPSPVVVGT 1075
```

RESULT 7
PCT-US03-19834-2
; Sequence 2, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; TITLE OF INVENTION: METHODS FOR TREATING HYPERLIPIDEMIA
; FILE REFERENCE: IOWA:045W0
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-2

Query Match 100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.1e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFPTSPVWVGT 20
Db 499 AKSVCGPVYCFPTSPVWVGT 518

RESULT 8
US-10-296-734-406
; Sequence 406, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la consensus polypeptide
US-10-296-734-406

Query Match 100.0%; Score 20; DB 6; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.1e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFPTSPVWVGT 20
Db 499 AKSVCGPVYCFPTSPVWVGT 518

RESULT 9
US-10-296-734-810
; Sequence 810, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734

; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 810
; LENGTH: 5985
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la savine
US-10-296-734-810

Query Match 100.0%; Score 20; DB 6; Length 5985;
Best Local Similarity 100.0%; Pred. No. 1.7e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFPTSPVWVGT 20
Db 3066 AKSVCGPVYCFPTSPVWVGT 3085

RESULT 10
US-10-685-435-22
; Sequence 22, Application US/10685435
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Multi-mer peptides derived from Hepatitis C Virus
; TITLE OF INVENTION: envelope proteins for diagnostic use and vaccination
; TITLE OF INVENTION: purposes.
; FILE REFERENCE: PCT98.75.HCV30
; CURRENT APPLICATION NUMBER: US/10/685,435
; CURRENT FILING DATE: 2003-10-16
; PRIOR APPLICATION NUMBER: US/09/566,266B
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 97870179.5
; PRIOR FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: patenting Ver. 2.1
; SEQ ID NO 22
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-685-435-22

Query Match 85.0%; Score 17; DB 6; Length 31;
Best Local Similarity 100.0%; Pred. No. 4e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 VCGPVYCFPTSPVWVGT 20
Db 3 VCGPVYCFPTSPVWVGT 19

RESULT 11
US-10-664-391-9
; Sequence 9, Application US/10664391
; GENERAL INFORMATION:
; APPLICANT: Donnelly, John J.
; APPLICANT: Liu, Margaret A.
; APPLICANT: Shiver, John W.
; APPLICANT: Fu, Tong-Ming
; TITLE OF INVENTION: SYNTHETIC HEPATITIS C GENES
; FILE REFERENCE: 19732YPCA
; CURRENT APPLICATION NUMBER: US/10/664,391
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: PCT/US97/09884
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/033,534
; PRIOR FILING DATE: 1996-12-20
; PRIOR APPLICATION NUMBER: 60/020,494
; PRIOR FILING DATE: 1996-06-11
; NUMBER OF SEQ ID NOS: 25

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-10-664-391-9

Query Match      85.0%; Score 17; DB 6; Length 347;
Best Local Similarity 100.0%; Pred. No. 1.9e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 VCGPVYCFPTSPVVVGT 20
Db      120 VCGPVYCFPTSPVVVGT 136

RESULT 12
US-10-664-391-11
; Sequence 11, Application US/10664391
; GENERAL INFORMATION:
; APPLICANT: Donnelly, John J.
; APPLICANT: Liu, Margaret A.
; APPLICANT: Shiver, John W.
; APPLICANT: Fu, Tong-Ming
; TITLE OF INVENTION: SYNTHETIC HEPATITIS C GENES
; FILE REFERENCE: 197321PCA
; CURRENT APPLICATION NUMBER: US/10/664,391
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: PCT/US97/09884
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/033,534
; PRIOR FILING DATE: 1996-12-20
; PRIOR APPLICATION NUMBER: 60/020,494
; PRIOR FILING DATE: 1996-06-11
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-10-664-391-11

Query Match      85.0%; Score 17; DB 6; Length 539;
Best Local Similarity 100.0%; Pred. No. 2.6e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 VCGPVYCFPTSPVVVGT 20
Db      312 VCGPVYCFPTSPVVVGT 328

RESULT 13
PCT-US03-20409-3
; Sequence 3, Application PC/TUS0320409
; GENERAL INFORMATION:
; APPLICANT: Merck & Co., Inc.
; TITLE OF INVENTION: HEPATITIS C VIRUS PARTICLE FORMATION
; FILE REFERENCE: 21123 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/20409
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/393,167
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1026
; TYPE: PRT
; ORGANISM: HCV
PCT-US03-20409-3

Query Match      85.0%; Score 17; DB 1; Length 1026;
Best Local Similarity 100.0%; Pred. No. 4e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-10-664-391-9

Query Match      85.0%; Score 17; DB 6; Length 347;
Best Local Similarity 100.0%; Pred. No. 1.9e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 VCGPVYCFPTSPVVVGT 20
Db      120 VCGPVYCFPTSPVVVGT 136

RESULT 14
PCT-US03-20322-211
; Sequence 211, Application PC/TUS0320322
; GENERAL INFORMATION:
; APPLICANT: Dana-Farber Cancer Institute, Inc., et al.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; FILE REFERENCE: DEN-043PC
; CURRENT APPLICATION NUMBER: PCT/US03/20322
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 211
; LENGTH: 2280
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
PCT-US03-20322-211

Query Match      85.0%; Score 17; DB 1; Length 2280;
Best Local Similarity 100.0%; Pred. No. 6.7e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 VCGPVYCFPTSPVVVGT 20
Db      502 VCGPVYCFPTSPVVVGT 518

RESULT 15
US-10-009-002-5
; Sequence 5, Application US/10009002
; GENERAL INFORMATION:
; APPLICANT: Bukh, Jens
; APPLICANT: Purcell, Robert
; APPLICANT: Yanagi, Masayuki
; APPLICANT: Emerson, Suzanne
; TITLE OF INVENTION: Infectious cDNA Clone of GB virus B and
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: NIH257.001NP
; CURRENT APPLICATION NUMBER: US/10/009,002
; CURRENT FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: PCT/US00/15293
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: US 60/137,694
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 3033
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-009-002-5

Query Match      85.0%; Score 17; DB 6; Length 3033;
Best Local Similarity 100.0%; Pred. No. 8.1e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 VCGPVYCFPTSPVVVGT 20
Db      504 VCGPVYCFPTSPVVVGT 520

Search completed: November 21, 2003, 22:12:56
Job time : 9.55 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:46:27 ; Search time 9.25 seconds
(without alignments)
207.932 Million cell updates/sec

Title: US-09-973-025-80

Perfect score: 20

Sequence: 1 AKSVCGPYCFTSPVVGVT 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 9616862 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR_76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	3011	1 GNMVCH	genome polyprotein
2	20	100.0	3011	1 GNMVCH	genome polyprotein
3	20	100.0	3011	1 S40770	genome polyprotein
4	18	90.0	3033	1 JQ1303	genome polyprotein
5	17	85.0	234	2 S32742	genome polyprotein
6	17	85.0	235	2 S32747	genome polyprotein
7	17	85.0	237	2 S32744	genome polyprotein
8	17	85.0	415	2 PC4407	envelope protein -
9	17	85.0	520	2 JQ1925	polyprotein - hepa
10	17	85.0	523	2 JQ1926	polyprotein - hepa
11	17	85.0	640	2 JQ1584	genome polyprotein
12	17	85.0	716	2 JQ1366	polyprotein - hepa
13	17	85.0	782	2 S19876	genome polyprotein
14	17	85.0	782	2 S18031	genome polyprotein
15	17	85.0	782	2 S18032	genome polyprotein
16	17	85.0	782	2 S19875	genome polyprotein
17	17	85.0	787	2 PN0677	hypothetical prote
18	17	85.0	876	2 PC2219	polypeptide - hepa
19	17	85.0	3010	1 GNMVTC	genome polyprotein
20	17	85.0	3010	1 GNMVCJ	genome polyprotein
21	17	85.0	3010	1 A45573	genome polyprotein
22	17	85.0	3010	1 S18030	genome polyprotein
23	17	85.0	3010	1 GNMVTW	genome polyprotein
24	17	85.0	3014	1 JCS620	genome polyprotein
25	17	85.0	3033	1 GNMVJ8	genome polyprotein
26	15	75.0	513	2 A44150	structural protein
27	14	70.0	550	2 JH0711	genome polyprotein
28	12	60.0	513	2 PC1284	genome polyprotein
29	8	40.0	138	2 S24081	envelope protein -

30	8	40.0	350	2 S35631	genome polyprotein
31	7	35.0	418	2 T47646	hypothetical prote
32	7	35.0	874	2 JQ0881	genome polyprotein
33	6	30.0	82	1 VLDK1	apovitellenin I -
34	6	30.0	95	1 KRDKF4	keratin B-4, feath
35	6	30.0	95	1 KRPYF4	keratin B-4, feath
36	6	30.0	98	1 KRCHF1	keratin I, feathe
37	6	30.0	98	1 KRCHF2	keratin II, feathe
38	6	30.0	98	2 S06807	keratin, feather (
39	6	30.0	98	2 S06806	keratin, feather (
40	6	30.0	98	2 S06808	keratin, feather (
41	6	30.0	102	1 KREUB	keratin, feather -
42	6	30.0	115	2 S06809	keratin - chicken
43	6	30.0	119	2 T49363	hypothetical prote
44	6	30.0	138	2 S24080	envelope protein -
45	6	30.0	168	2 E83437	hypothetical prote

ALIGNMENTS

RESULT 1

GNMVC3

genome polyprotein - hepatitis C virus (strain HCV-1)

N:Contains: capsid protein C; envelope protein M; hepatitisvirin (EC 3.4.21.98) (nonstructura

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C>Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 19-Jan-2001

C:Accession: A39166; PQ0403; PQ0404

R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Co;

Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991

A:Title: Genetic organization and diversity of the hepatitis C virus.

A:Reference number: A39166; MUID:91172826; PMID:1848704

A:Accession: A39166

A:Molecule type: mRNA

A:Residues: 1-3011 <CHO>

A:Cross-references: GB:MG2321; NID:g329873; PIDN:AAA45676.1; PID:g329874

J:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.I.

J. Gen. Virol. 73, 1131-1141, 1992

A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e

A:Reference number: PQ0393; MUID:92268871; PMID:1316939

A:Accession: PQ0403

A:Molecule type: Genomic RNA

A:Residues: 1577-1633 <CHA>

A:Cross-references: DDBJ:D10128

A:Experimental source: isolates E-b16

A:Accession: PQ0404

A>Status: preliminary

A:Molecule type: Genomic RNA

A:Residues: 1577-1633 <CH2>

A:Experimental source: isolates E-b17

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura

F:115/Product: capsid protein C #status predicted <CBC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitisvirin #status predicted <NS3>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,476,532,556,576,623,645,1213,1255,2041,2077,22

Query Match 100.0%; Score 20; DB 1; Length 3011;

Best Local Similarity 100.0%; Pred. No. 3.9e-12;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPYCFTSPVVGVT 20

|||||

Db 499 AKSVCGPVYCFTSPVVVGT 518

RESULT 2
GNWVCH
genome polyprotein - hepatitis C virus (strain H)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A36814; A41546
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: c
A:Reference number: A36814
A:Accession: A36814
A:Molecule type: genomic RNA
A:Residues: 1-3011 <INC>
A:Cross-references: GB:M67463; NID:G329737; PIDN:AAA45534.1; PID:G329738
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: compar
A:Reference number: A41546; MUID:92052256; PMID:1658800
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:730-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match 100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.9e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFTSPVVVGT 20
Db 499 AKSVCGPVYCFTSPVVVGT 518

RESULT 3
S40770
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: S40770; PC1285
R:Okamoto, H.
submitted to the EMBL Data Library, March 1992
A:Reference number: S40770
A:Accession: S40770
A:Molecule type: genomic RNA
A:Residues: 1-3011 <OKA>
A:Cross-references: EMBL:D10749; NID:G221586; PIDN:BAA01582.1; PID:G221587
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,
Jun. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; MUID:91013116; PMID:2170712
A:Accession: PC1285
A:Molecule type: genomic RNA
A:Residues: 1-513 <OK2>

A:Cross-references: GB:D00831; NID:G221511; PIDN:BAA00705.1; PID:G221512
A:Experimental source: isolate HC-J1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:730-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.9e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFTSPVVVGT 20
Db 499 AKSVCGPVYCFTSPVVVGT 518

RESULT 4
J01303
genome polyprotein - hepatitis C virus (isolate HC-J6)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
C:Accession: J01303
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.
J. Gen. Virol. 72, 2697-2704, 1991
A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a hum
A:Reference number: J01303; MUID:92044440; PMID:1658196
A:Accession: J01303
A:Molecule type: genomic RNA
A:Residues: 1-3033 <OKA>
A:Cross-references: GB:D00944; NID:G221650; PIDN:BAA00792.1; PID:G221651
A:Experimental source: isolate HC-J6 from a Japanese individual
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; trans
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:730-733/Product: nonstructural protein NS1 #status predicted <NS1>
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F:1011-1619/Product: hepatitis C virus genome polyprotein
F:1316-1321/Region: nucleotide-binding motif B
F:1320-1323/Region: DEXH motif
F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28

Query Match 90.0%; Score 18; DB 1; Length 3033;
Best Local Similarity 100.0%; Pred. No. 3.9e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 SVCGPVYCFTSPVVVGT 20
Db 503 SVCGPVYCFTSPVVVGT 520

RESULT 5
S32742
genome polyprotein - hepatitis C virus (isolate CR-1) (fragment)
N:Contains: envelope protein E2
C:Species: hepatitis C virus
A:Variety: isolate CR-1

C>Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S32742
R:Roggendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
submitted to the EMBL Data Library, February 1993
A:Description: Variability of the envelope regions of HCV in European isolates and its
A:Reference number: S32741
A:Accession: S32742
A:Molecule type: genomic RNA
A:Residues: 1-234 <ROG>
A:Cross-references: EMBL:X72979; NID:g296102; PIDN:CAA51484.1; PID:g296103
A:Experimental source: isolate Cr-1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; polyprotein
F:1-234/Product: envelope protein E2 #status predicted <MAT>

Query Match 85.0%; Score 17; DB 2; Length 234;
Best Local Similarity 100.0%; Pred. No. 5.1e-10; Mismatches 0; Indels 0; Gaps 0;
Matches 17; Conservative 0;

Qy 4 VCGPVYCFPTSPVVVGT 20
|||||
Db 133 VCGPVYCFPTSPVVVGT 149
|||||

RESULT 6
S32747
genome polyprotein - hepatitis C virus (isolate HU-1) (fragment)
N:Contains: envelope protein E2
C:Species: hepatitis C virus
A:Variety: isolate HU-1
C>Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S32747
R:Roggendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
submitted to the EMBL Data Library, February 1993
A:Description: Variability of the envelope regions of HCV in European isolates and its
A:Reference number: S32741
A:Accession: S32747
A:Molecule type: genomic RNA
A:Residues: 1-235 <ROG>
A:Cross-references: EMBL:X72977; NID:g296112; PIDN:CAA51482.1; PID:g296113
A:Experimental source: isolate HU-1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; polyprotein
F:1-235/Product: envelope protein E2 #status predicted <MAT>

Query Match 85.0%; Score 17; DB 2; Length 235;
Best Local Similarity 100.0%; Pred. No. 5.1e-10; Mismatches 0; Indels 0; Gaps 0;
Matches 17; Conservative 0;

Qy 4 VCGPVYCFPTSPVVVGT 20
|||||
Db 134 VCGPVYCFPTSPVVVGT 150
|||||

RESULT 7
S32744
genome polyprotein - hepatitis C virus (isolate EG-1) (fragment)
N:Contains: envelope protein E2
C:Species: hepatitis C virus
A:Variety: isolate EG-1
C>Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S32744
R:Roggendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
submitted to the EMBL Data Library, February 1993
A:Description: Variability of the envelope regions of HCV in European isolates and its
A:Reference number: S32741
A:Accession: S32744
A:Molecule type: genomic RNA
A:Residues: 1-237 <ROG>
A:Cross-references: EMBL:X72981
A:Experimental source: isolate EG-1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; polyprotein

F:1-237/Product: envelope protein E2 #status predicted <MAT>

Query Match 85.0%; Score 17; DB 2; Length 237;
Best Local Similarity 100.0%; Pred. No. 5.2e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 VCGPVYCFPTSPVVVGT 20
|||||
Db 136 VCGPVYCFPTSPVVVGT 152
|||||

RESULT 8
PC4407
envelope protein - hepatitis C virus (fragment)
C:Species: hepatitis C virus
C>Date: 10-Nov-1997 #sequence_revision 23-Jan-1998 #text_change 17-Nov-2000
C:Accession: PC4407
R:Li, G.; Yao, J.; Peng, W.
Chinese J. Virol. 13, 24-32, 1997
A:Title: Sequence of genomic region of hepatitis C virus envelope proteins from a Guang
A:Reference number: PC4407
A:Accession: PC4407
A:Molecule type: genomic RNA
A:Residues: 1-415 <LIA>
A:Note: the authors translated the codon ATA for residues 93 and 249 as Met
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein

Query Match 85.0%; Score 17; DB 2; Length 415;
Best Local Similarity 100.0%; Pred. No. 8.1e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 VCGPVYCFPTSPVVVGT 20
|||||
Db 337 VCGPVYCFPTSPVVVGT 353
|||||

RESULT 9
JQ1925
polyprotein - hepatitis C virus (isolate HCV-KP)
N:Contains: C protein; E1 protein; E2/NS1 protein
C:Species: hepatitis C virus
C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 17-Nov-2000
C:Accession: JQ1925
R:Abe, K.; Inchauspe, G.; Fujisawa, K.
J. Gen. Virol. 73, 2725-2729, 1992
A:Title: Genomic characterization and mutation rate of hepatitis C virus isolated from
A:Reference number: JQ1925; MUID:93019030; PMID:1383400
A:Accession: JQ1925
A:Molecule type: mRNA
A:Residues: 1-520 <ABE>
A:Cross-references: DDBJ:D10687; NID:g221544; PIDN:BAA01529.1; PID:g221545
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein; transmembrane protein
F:1-191/Product: C protein #status predicted <CPR>
F:192-383/Product: E1 protein #status predicted <E1R>
F:384-520/Product: E2/NS1 protein #status predicted <E2P>

Query Match 85.0%; Score 17; DB 2; Length 520;
Best Local Similarity 100.0%; Pred. No. 9.7e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 VCGPVYCFPTSPVVVGT 20
|||||
Db 503 VCGPVYCFPTSPVVVGT 519
|||||

RESULT 10
JQ1926
polyprotein - hepatitis C virus (isolate HCV-476)
N:Contains: C protein; E1 protein; E2/NS1 protein
C:Species: hepatitis C virus
C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 17-Nov-2000

C;Accession: J01926
 R;Abe, K.; Inchauspe, G.; Fujisawa, K.
 J. Gen. Virol. 73, 2725-2729, 1992
 A;Title: Genomic characterization and mutation rate of hepatitis C virus isolated from a
 A;Reference number: JQ1925; PMID:1383400
 A;Accession: J01926
 A;Molecule type: mRNA
 A;Residues: 1-523 <ABE>
 A;Cross-references: DDBJ:D10687
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: polyprotein
 F;1-191/Product: C protein #status predicted <CPR>
 F;192-383/Product: E1 protein #status predicted <E1R>
 F;384-523/Product: E2/NS1 protein #status predicted <E2P>

Query Match 85.0%; Score 17; DB 2; Length 523;
 Best Local Similarity 100.0%; Pred. No. 9.7e-10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 VCGPVYCFPTSPVVVGT 20
 Db 503 VCGPVYCFPTSPVVVGT 519

RESULT 11
 JQ1584
 genome polyprotein - hepatitis C virus (strain U.K.) (fragment)
 N;Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural pro
 C;Species: hepatitis C virus
 C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
 C;Accession: JQ1584
 R;Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.
 J. Gen. Virol. 73, 1521-1525, 1992
 A;Title: Cloning and sequencing of the structural region and expression of putative core
 A;Reference number: JQ1584; PMID:92300349; PMID:1318944
 A;Accession: JQ1584
 A;Molecule type: genomic RNA
 A;Residues: 1-640 <KUM>
 A;Cross-references: GB:X84079; NID:9643119; PIDN:CAA58888.1; PID:9643120
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; polypro
 F;1-191/Product: core protein C #status predicted <CPC>
 F;192-389/Product: envelope protein E1 #status predicted <EE1>
 F;390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predicted <
 F;196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn) (cova

Query Match 85.0%; Score 17; DB 2; Length 640;
 Best Local Similarity 100.0%; Pred. No. 1.1e-09;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 VCGPVYCFPTSPVVVGT 20
 Db 502 VCGPVYCFPTSPVVVGT 518

RESULT 12
 JQ1366
 polyprotein - hepatitis C virus (French isolate) (fragments)
 C;Species: hepatitis C virus
 C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
 C;Accession: JQ1366
 R;Kremsdorf, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.
 J. Gen. Virol. 72, 2557-2561, 1991
 A;Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication
 A;Reference number: JQ1366; PMID:92013977; PMID:1655961
 A;Accession: JQ1366
 A;Molecule type: genomic RNA
 A;Residues: 1-716 <KRE>
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: glycoprotein; polyprotein
 F;84,90,97,115,143,199,223,290,312/Binding site: carbohydrate (Asn) (covalent) #stat

Query Match 85.0%; Score 17; DB 2; Length 716;

Best Local Similarity 100.0%; Pred. No. 1.3e-09;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 VCGPVYCFPTSPVVVGT 20
 Db 169 VCGPVYCFPTSPVVVGT 185

RESULT 13
 S19876
 genome polyprotein - hepatitis C virus (isolate JK5) (fragment)
 N;Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
 C;Species: hepatitis C virus
 A;Variety: isolate JK5
 C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
 C;Accession: S19876
 R;Honda, M.; Kaneko, S.; Maesashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A;Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
 A;Reference number: S18029
 A;Accession: S19876
 A;Molecule type: genomic RNA
 A;Residues: 1-782 <HON>
 A;Cross-references: EMBL:X61595; NID:g59486; PIDN:CAA43792.1; PID:g59487
 A;Experimental source: isolate JK5
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: capsid protein; core protein; core protein; glycoprotein; glycoprotein; nonstructural
 F;1-191/Product: core protein #status predicted <MARI>
 F;192-383/Product: envelope protein 1 #status predicted <MAT2>
 F;384-733/Product: NS1/E2 protein #status predicted <MAT3>
 F;734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 85.0%; Score 17; DB 2; Length 782;
 Best Local Similarity 100.0%; Pred. No. 1.3e-09;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 VCGPVYCFPTSPVVVGT 20
 Db 502 VCGPVYCFPTSPVVVGT 518

RESULT 14
 S18031
 genome polyprotein - hepatitis C virus (isolate JK2) (fragment)
 N;Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
 C;Species: hepatitis C virus
 A;Variety: isolate JK2
 C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
 C;Accession: S18031
 R;Honda, M.; Kaneko, S.; Maesashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A;Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
 A;Reference number: S18029
 A;Accession: S18031
 A;Molecule type: genomic RNA
 A;Residues: 1-782 <HON>
 A;Cross-references: EMBL:X61593
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: capsid protein; core protein; glycoprotein; glycoprotein; nonstructural
 F;1-191/Product: core protein #status predicted <MAT1>
 F;192-383/Product: core protein #status predicted <MAT2>
 F;384-733/Product: NS1/E2 protein #status predicted <MAT3>
 F;734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 85.0%; Score 17; DB 2; Length 782;
 Best Local Similarity 100.0%; Pred. No. 1.3e-09;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 VCGPVYCFPTSPVVVGT 20
 Db 502 VCGPVYCFPTSPVVVGT 518


```

RESULT 15
S18032
genome polyprotein - hepatitis C virus (isolate JK4) (fragment)
N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C:Species: hepatitis C virus
A:Variety: isolate JK4
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C:Accession: S18032
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
A:Reference number: S18029
A:Accession: S18032
A:Molecule type: genomic RNA
A:Residues: 1-782 <HON>
A:Cross-references: EMBL:X51594
A:Experimental source: isolate JK4
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F:1-191/Product: core protein #status predicted <MAT1>
F:192-383/Product: envelope protein 1 #status predicted <MAT2>
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match      85.0%; Score 17; DB 2; Length 782;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 VCGPVYCFPTSPVWVGT 20
      |||||
DB     502 VCGPVYCFPTSPVWVGT 518
  
```

Search completed: November 21, 2003, 21:11:32
 Job time : 10.25 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:52 ; Search time 4.85 Seconds
(without alignments)
193.925 Million cell updates/sec

Title: US-09-973-025-80

Perfect score: 20

Sequence: 1 AKSVGPVYCFTPSPVVGT 20

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query	ID	Description
1	20	100.0	3011	1	POLG_HCV1
2	20	100.0	3011	1	POLG_HCVH
3	18	90.0	737	1	POLG_HCVJ5
4	18	90.0	3033	1	POLG_HCVJ6
5	17	85.0	520	1	POLG_HCVH4
6	17	85.0	520	1	POLG_HCVHK
7	17	85.0	737	1	POLG_HCVJ7
8	17	85.0	3010	1	POLG_HCVBK
9	17	85.0	3010	1	POLG_HCVJA
10	17	85.0	3010	1	POLG_HCVJT
11	17	85.0	3010	1	POLG_HCVTW
12	17	85.0	3033	1	POLG_HCVJ8
13	12	60.0	513	1	POLG_HCVJ2
14	6	30.0	82	1	APV1_ANAPL
15	6	30.0	95	1	KRFT_ANAPL
16	6	30.0	95	1	KRFT_COLLI
17	6	30.0	97	1	KRFA_CHICK
18	6	30.0	97	1	KRFB_CATAU
19	6	30.0	97	1	KRFB_CHICK
20	6	30.0	97	1	KRFB_MYCAM
21	6	30.0	97	1	KRFB_CHICK
22	6	30.0	97	1	KRFD_CHICK
23	6	30.0	100	1	KRFL_COLLI
24	6	30.0	100	1	KRFL_COLLI
25	6	30.0	102	1	KRFT_DRONO
26	6	30.0	258	1	UPKA_BOVIN
27	6	30.0	258	1	UPKA_HUMAN
28	6	30.0	297	1	DSR3_HUMAN
29	6	30.0	369	1	YCBX_ECOLI
30	6	30.0	394	1	KICE_MOUSE
31	6	30.0	394	1	KICE_RAT
32	6	30.0	395	1	KICE_HUMAN
33	6	30.0	401	1	HEM1_RHOCA

34 6 30.0 407 1 HEM1_RHOSH Q04512 rhodobacter
35 6 30.0 456 1 KICH_HUMAN P35790 homo sapien
36 6 30.0 496 1 DFNS_HUMAN O60443 homo sapien
37 6 30.0 501 1 ALG8_SCHPO Q10479 schizosacch
38 6 30.0 518 1 AGM1_SCHPO Q09687 schizosacch
39 6 30.0 525 1 BCHB_RHOCA P26163 rhodobacter
40 6 30.0 579 1 GPC2_RAT P51653 rattus norv
41 6 30.0 587 1 UVSW_BPT4 P20703 bacterioph
42 6 30.0 652 1 PICA_HUMAN Q13492 homo sapien
43 6 30.0 660 1 PAR3_ARATH O64380 arabidopsis
44 6 30.0 674 1 CIN_DROME P39205 drosophila
45 6 30.0 1008 1 VGLM_DUK P09613 ukuniemi v

ALIGNMENTS

RESULT 1

POLG_HCV1 STANDARD; PRT; 3011 AA.
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate 1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA Bradley D.W., Kuo G., Houghton M.;
RT "Genetic organization and diversity of the hepatitis C virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA)(N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: OF
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M62321; AAA45676.1; -;
CC PIR; A39166; GNMVVC3.
CC PDB; 1A1V; 16-FEB-99.
CC PDB; 1HEI; 25-NOV-98.
CC MEROPS; S29.001; -;
CC MEROPS; U39.001; -;
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002522; HCV_capsid.

DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR004109; HCV NS3.
 DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002868; HCV NS5a.
 DR InterPro: IPR002166; HCV NS5b.
 DR InterPro: IPR001650; HCV NS5c.
 DR InterPro: IPR007095; RNA pol DS P5.
 DR InterPro: IPR007094; RNA pol PSV1.
 DR Pfam: PF01543; HCV capsid; 1.
 DR Pfam: PF01542; HCV core; 1.
 DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR Pfam: PF01538; HCV NS2; 1.
 DR Pfam: PF02907; HCV NS3; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF01001; HCV NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral RdRp; 1.
 DR ProDom: PD186062; HCV NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3011
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 476 476
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2364 2364
 FT CARBOHYD 2789 2789
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;
 Query Match 100.0%; Score 20; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 6e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AKSVCGPVYCFPTSPVVGVT 20

Db 499 AKSVCGPVYCFPTSPVVGVT 518
 RESULT 2
 POLG_HCVH
 ID POLG_HCVH STANDARD; PRT; 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, last sequence update)
 DT 15-SEP-2003 (Rel. 42, last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA, positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=1658800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154322; PubMed=9493270;
 RA Kim J.U., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding";
 RL Structure 6:89-100(1998).
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position. Cys or Thr in Pi and Ser or Ala in Pi'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY.
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M67463; AAA45534.1; -
DR PIR; A36814; GNWVCH.
DR PDB; 1HEI; 25-NOV-98.
DR PDB; 1AIV; 16-FEB-99.
DR PDB; 1AIR; 17-JUN-98.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR TRANSFAC; T04155; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA pol DS ps.
DR InterPro; IPR007094; RNA pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase C; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
FT CHAIN 192 383 CAPSID PROTEIN C.
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
FT CHAIN 810 1026 PROTEIN P7.
FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5A.
FT CHAIN 3011 369 NONSTRUCTURAL PROTEIN NS5B.
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH_BOX.
FT CARBOHYD 136 136 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).

FT STRAND 1224 1226
FT TURN 1232 1233
FT TURN 1236 1238
FT HELIX 1239 1246
FT TURN 1247 1248
FT STRAND 1251 1255
FT HELIX 1258 1271
FT TURN 1272 1272
FT STRAND 1277 1280
FT TURN 1281 1282
FT STRAND 1283 1285
FT STRAND 1291 1295
FT HELIX 1296 1301
FT TURN 1302 1303
FT STRAND 1312 1316
FT TURN 1317 1319
FT HELIX 1323 1335
FT TURN 1336 1340
FT STRAND 1343 1347
FT TURN 1352 1353
FT STRAND 1360 1361
FT STRAND 1362 1366
FT STRAND 1368 1368
FT STRAND 1373 1375
FT TURN 1376 1377
FT STRAND 1378 1380
FT HELIX 1382 1385
FT STRAND 1389 1393
FT HELIX 1397 1409
FT TURN 1410 1411
FT STRAND 1414 1417
FT TURN 1419 1420
FT STRAND 1432 1436
FT TURN 1438 1439
FT STRAND 1450 1453
FT STRAND 1456 1463
FT STRAND 1471 1478
FT STRAND 1480 1480
FT HELIX 1481 1488
FT TURN 1489 1490
FT STRAND 1497 1501
FT STRAND 1507 1507
FT STRAND 1511 1511
FT HELIX 1514 1527
FT STRAND 1532 1544
FT STRAND 1550 1550
FT HELIX 1555 1564
FT TURN 1570 1578
FT TURN 1579 1580
FT HELIX 1584 1597
FT TURN 1598 1598
FT HELIX 1606 1611
FT TURN 1614 1618
FT STRAND 1622 1623
FT STRAND 1627 1627
FT STRAND 1635 1636
FT HELIX 1640 1652
SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKSVCGPVYCFPSPVVVG 20
Db 499 AKSVCGPVYCFPSPVVVG 518

RESULT 3
POLG_HCVJ5
ID POLG_HCVJ5 STANDARD; PRT; 737 AA.
AC P27960;
DT 01-AUG-1992 (Rel. 23, Created)

DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR007095; RNA pol DS PS.
 DR InterPro; IPR007094; RNA pol PSVr.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; Helicase C; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR Polyprotein; Glycoprotein; Transferrase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolyase; Serine protease.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 1087
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2811 2811
 SQ SEQUENCE 3033 AA; 329165 MW; F957P5C1A273BE9E CRC64;

Query Match 90.0%; Score 18; DB 1; Length 3033;
 Best Local Similarity 100.0%; Pred. No. 7e-11;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 SVCGPVYCFTPSPVVVGT 20
 |||||
 Db 503 SVCGPVYCFTPSPVVVGT 520

RESULT 5
 POLG_HCVH4
 ID POLG_HCVH4 STANDARD; PRT; 520 AA.
 AC Q01404;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);

DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 (GP68) (GP70) (NS1) (Fragment).
 OS Hepatitis C virus (isolate HCV-476) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31643;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93019030; PubMed=1383400;
 RA Abe K., Inchauspe G., Fujisawa K.;
 RT "Genomic characterization and mutation rate of hepatitis C virus
 isolated from a patient who contracted hepatitis during an epidemic
 of non-A, non-B hepatitis in Japan.";
 RL J. Gen. Virol. 73:2725-2729 (1992).
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS;
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D10688; BAA01530.1; -;
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 >520
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 418 418
 FT CARBOHYD 424 424
 FT CARBOHYD 431 431
 FT CARBOHYD 449 449
 FT NON_TER 520 520
 SQ SEQUENCE 520 AA; 56499 MW; AA135246CF20D525 CRC64;

Query Match 85.0%; Score 17; DB 1; Length 520;
 Best Local Similarity 100.0%; Pred. No. 1.7e-10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VCGPVYCFTPSPVVVGT 20
 |||||
 Db 508 VCGPVYCFTPSPVVVGT 519

RESULT 6
 POLG_HCVH4
 ID POLG_HCVH4 STANDARD; PRT; 520 AA.
 AC Q01403;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

```
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (Fragment).
OS Hepatitis C virus (isolate HCV-KF) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31644;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=93019030; PubMed=1383400;
RA Abe K., Inchauspe G., Fujisawa K.;
RT "Genomic characterization and mutation rate of hepatitis C virus
RT isolated from a patient who contracted hepatitis during an epidemic
RT of non-A, non-B hepatitis in Japan.";
RL J. Gen. Virol. 73:2725-2729(1992).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D10687; BAA01529.1; -.
DR PIR; JQ1925; JQ1925.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR PolyProtein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 >520
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 418 418
FT CARBOHYD 424 424
FT CARBOHYD 431 431
FT CARBOHYD 449 449
FT NON_TER 520 520
SQ SEQUENCE 520 AA; 56476 MW; 1D2BD0A6FF27349B CRC64;

Query Match 85.0%; Score 17; DB 1; Length 520;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VCGPVYCFTPSPVVGVT 20
Db 503 VCGPVYCFTPSPVVGVT 519
|||||
|||||

RESULT 7
POLG_HCVJ7 STANDARD; PRT; 737 AA.
AC P27961;
```

```
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
DE protein (Envelope protein M); Major envelope protein E; Nonstructural
DE proteins NS1 and NS2] (Fragment).
OS Hepatitis C virus (isolate HC-J7) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11114;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D10077; BAA00971.1; -.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR PolyProtein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 >737
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 299 299
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT NON_TER 737 737
SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122F2 CRC64;
```


Query Match 85.0%; Score 17; DB 1; Length 737;
Best Local Similarity 100.0%; Pred. No. 2.3e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VCGPVYCFPTSPVVVGT 20
DB 504 VCGPVYCFPTSPVVVGT 520

RESULT 8
POLG_HCVBK STANDARD; PRT; 3010 AA.

AC P26663;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)).
OS Hepatitis C virus (isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11105;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91140698; PubMed=1847440;
RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA Onishi E., Andoh T., Yoshida I., Okayama H.;
RT "Structure and organization of the hepatitis C virus genome isolated
RT from human carriers";
RL J. Virol. 65:1105-1113 (1991).
RN [2]
RP SEQUENCE OF 1487-1500.
RX MEDLINE=96235224; PubMed=8647104;
RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetky L.;
RT "Non-structural protein 3 of hepatitis C virus inhibits
RT phosphorylation mediated by cAMP-dependent protein kinase";
RL Eur. J. Biochem. 237:611-618 (1996).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX MEDLINE=97015098; PubMed=8861916;
RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
RA Moonaw E.W., Adachi T., Hostomsky Z.;
RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT trypsin-like fold and a structural zinc binding site";
RL Cell 87:331-342 (1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX MEDLINE=98227846; PubMed=9568891;
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RT virus: a 2.2-A resolution structure in a hexagonal crystal form";
RL Protein Sci. 7:837-847 (1998).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M58335; AAA72945.1; -;
DR PIR; A38465; GNWVTC
DR PDB; 1AIQ; 25-MAR-98.
DR PDB; 1UXP; 14-JAN-98.
DR PDB; 1NS3; 08-APR-98.
DR PDB; 1C2P; 15-NOV-00.
DR PDB; 1CSJ; 08-NOV-99.
DR PDB; 1GX5; 09-APR-02.
DR PDB; 1GX6; 10-APR-02.
DR PDB; 1QUV; 26-JUN-00.
DR PDB; 8OHM; 20-APR-99.
DR MEROPS; S29.001; -;
DR MEROPS; U39.001; -;
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR007095; RNA pol PS.
DR InterPro; IPR007094; RNA pol PSvir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00998; Viral RdRp; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXdc; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CELLULAR AMINOPEPTIDASE
FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM.
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1031 1035 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1039 1047 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1050 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1059 1063 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1068 1074 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1075 1076 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1077 1081 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1082 1085 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1086 1087 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1090 1092 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1093 1094 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1095 1097 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1101 1103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1104 1107 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1108 1112 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1120 1120 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1122 1122 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1129 1133 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1135 1136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1139 1144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1149 1157 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1158 1161 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1162 1163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1165 1166 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1168 1171 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1172 1174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1175 1186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1187 1188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1189 1197 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1198 1202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1203 1204 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1680 1688 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECFFDP9C CRC64;

Query Match
Best Local Similarity 100.0%; Score 17; DB 1; Length 3010;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VCGPVYCFPPSPVVGVT 20
Db 502 VCGPVYCFPPSPVVGVT 518

RESULT 9
POLG HCVJA
ID POLG HCVJA STANDARD; PRT; 3010 AA.
AC P26662;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

Hepacivirus.
NCBI_TaxID=11116;
[1]
SEQUENCE FROM N.A.
MEDLINE=91088550; PubMed=2175903;
Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
Sugimura T., Shimotohno K.;
"Molecular cloning of the human hepatitis C virus genome from
Japanese patients with non-A, non-B hepatitis";
Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
[2]
DISCUSSION OF SEQUENCE.
MEDLINE=91192160; PubMed=1849488;
Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraio K.,
Ohkoshi S., Shimotohno K.;
"Molecular structure of the Japanese hepatitis C viral genome.";
FEBS Lett. 280:325-328(1991).
-!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
-!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
precursor polyprotein, commonly with Asp or Glu in the P6
position, Cys or Thr in P1 and Ser or Ala in P1'.
-!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
{RNA} (N).
-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA.
-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; D90208; BAA14233.1; -.
PIR; A39253; GNMVCJ.
HSSP; P26663; 1JXP.
MEROPS; S29.001; -.
MEROPS; U39.001; -.
InterPro; IPR001410; DEAD.
InterPro; IPR002522; HCV_capsid.
InterPro; IPR002521; HCV core.
InterPro; IPR002519; HCV env.
InterPro; IPR002531; HCV NS1.
InterPro; IPR002518; HCV NS2.
InterPro; IPR004109; HCV NS3.
InterPro; IPR000745; HCV NS4a.
InterPro; IPR001490; HCV NS4b.
InterPro; IPR002868; HCV NS5a.
InterPro; IPR002166; HCV RdRp.
InterPro; IPR001650; Helicase_C.
InterPro; IPR007095; RNA_pol_DS_PS.
InterPro; IPR007094; RNA_pol_PSVir.
Pfam; PF01543; HCV_capsid; 1.
Pfam; PF01542; HCV_core; 1.
Pfam; PF01539; HCV_env; 1.
Pfam; PF01560; HCV_NS1; 1.
Pfam; PF01538; HCV_NS2; 1.
Pfam; PF02907; HCV_NS3; 1.
Pfam; PF01006; HCV_NS4a; 1.
Pfam; PF01001; HCV_NS4b; 1.
Pfam; PF01506; HCV_NS5a; 1.
Pfam; PF00271; helicase_C; 1.
Pfam; PF00998; Viral_RdRp; 1.
Pfam; PF0186062; HCV_NS1; 1.
ProDom; PDI86062; HCV_NS1; 1.
SMART; SM00487; DEXoc; 1.
PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW

```



```

FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 1316 1319 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;

Query Match 85.0%; Score 17; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 7.5e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VCGPVYCFPTSPVWVGT 20
Db 502 VCGPVYCFPTSPVWVGT 518
|||||
|||||

RESULT 11
POLG_HCVTW STANDARD; PRT; 3010 AA.
AC P29946;
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Taiwan) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31645;
RN [1]
RP MEDLINE=92230206; PubMed=1314449;
RX Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
RA "The Taiwanese hepatitis C virus genome: sequence determination and
RT mapping the 5' terminus of viral genomic and antigenomic RNA.";
RL Virology 188:102-113(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

```

```

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/ebis/announce/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; M84754; -, NOT_ANNOTATED_CDS.
CC PIR; A40244; GNMVTV.
CC PDB; 1N64; 25-FEB-03.
CC PDB; 1NS3; 08-APR-98.
CC MEROPS; S29.001; -.
CC MEROPS; U39.001; -.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002522; HCV_capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC InterPro; IPR002531; HCV_NS1.
CC InterPro; IPR002518; HCV_NS2.
CC InterPro; IPR004109; HCV_NS3.
CC InterPro; IPR000745; HCV_NS4A.
CC InterPro; IPR001490; HCV_NS4B.
CC InterPro; IPR002868; HCV_NS5A.
CC InterPro; IPR002166; HCV_RdRP.
CC InterPro; IPR007095; RNA_pol_DS_PS.
CC InterPro; IPR007094; RNA_pol_PSVir.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NS1; 1.
CC Pfam; PF01538; HCV_NS2; 1.
CC Pfam; PF02907; HCV_NS3; 1.
CC Pfam; PF01006; HCV_NS4A; 1.
CC Pfam; PF01001; HCV_NS4B; 1.
CC Pfam; PF01506; HCV_NS5A; 1.
CC Pfam; PF00271; helicase_C; 1.
CC Pfam; PF00998; Viral_RdRP; 1.
CC ProDom; PD186062; HCV_NS1; 1.
CC SMART; SM00487; DEXBC; 1.
CC Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1
FT CHAIN 1 115 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 116 191 CELLULAR AMINOPEPTIDASE.
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 3010 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; AAD267D55CDFE215 CRC64;
 Query Match 85.0%; Score 17; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 7.5e-10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 VCGPVYCTPSPVVGVT 20
 DB 502 VCGPVYCTPSPVVGVT 518
 RESULT 12
 POLG_HCVJ8 STANDARD; PRT; 3033 AA.
 AC P26651;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P66); Nonstructural protein NS5A (P56); Nonstructural protein
 DE Hepatitis C virus (isolate HC-J8) (HCV).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11115;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
 RA Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes".
 RL Virology 188:331-341(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY 629.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; D10988; BAA01761.1; --
 CC FIR; A40250; GNMVJ8.
 CC HSSP; P27958; 1HEI.
 CC MEROPS; S29.001; --

DR MEROPS; U39.001; --
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV NS5a.
 DR InterPro; IPR007095; RNA pol DS PS.
 DR InterPro; IPR007094; RNA pol PSvir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00998; Viral_RDRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT SITE 1087 1087
 FT ACT SITE 1111 1111
 FT ACT SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2359 2359
 FT CARBOHYD 2811 2811
 SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;
 Query Match 85.0%; Score 17; DB 1; Length 3033;
 Best Local Similarity 100.0%; Pred. No. 7.6e-10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 VCGPVYCTPSPVVGVT 20

```

Db          504 VCGPVYCFPTSPVWGT 520
|||||
RESULT 13
POLG_HCVJ2
ID POLG_HCVJ2 STANDARD; PRT; 513 AA.
AC P27959;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (Fragment).
OS Hepatitis C virus (isolate HC-J2) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11111;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Teada F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes."
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
-----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: D10074; BAA00968.1; -
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR ProDom: PD186062; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
Transmembrane; Nonstructural protein.
FT INIT_MET 1
FT CHAIN 1
FT CHAIN 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 >513
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 448
FT CARBOHYD 448 448
FT NON_TER 513

```

```

SQ SEQUENCE 513 AA; 55704 MW; 943F31E3514CDEF3 CRC64;
Query Match 60.0%; Score 12; DB 1; Length 513;
Best Local Similarity 100.0%; Pred. No. 2.5e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 VCGPVYCFPTSP 15
DB 502 VCGPVYCFPTSP 513
|||||
RESULT 14
APV1_ANAPL
ID APV1_ANAPL STANDARD; PRT; 82 AA.
AC P02658;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Apovitelin I.
OS Anas platyrhynchos (Domestic duck).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anas.
OX NCBI_TaxID=8839;
RN [1]
RP SEQUENCE.
RX MEDLINE=77116126; PubMed=838046;
RA Inglis A.S., Burley R.W.;
RT "Determination of the amino acid sequence of apovitellenin I from
RT duck's egg yolk using an improved sequenator procedure: a comparison
RT with other avian species."
RL FEBS Lett. 73:33-37(1977).
CC -!- FUNCTION: PROTEIN COMPONENT OF THE VERY LOW DENSITY LIPOPROTEIN
CC (VLDL) OF EGG-LAYING FEMALES. POTENT LIPOPROTEIN LIPASE INHIBITOR,
CC PREVENTING THE LOSS OF TRIGLYCERIDES FROM VLDL ON THEIR WAY FROM
CC THE LIVER TO THE GROWING OOCYTES.
CC -!- SUBUNIT: Monomer (Potential).
CC -!- TISSUE SPECIFICITY: FOUND IN EGG YOLK AND IN PLASMA.
DR PIR; A03102; VLDK1.
KW Storage protein; VLDL.
SQ SEQUENCE 82 AA; 9492 MW; 3C80C929F6B3C93A CRC64;
Query Match 30.0%; Score 6; DB 1; Length 82;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 PVVWGT 20
DB 44 PVVWGT 49
|||||
RESULT 15
KRFT_ANAPL
ID KRFT_ANAPL STANDARD; PRT; 95 AA.
AC P08335;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Feather keratin B-4 (F-ker).
OS Anas platyrhynchos (Domestic duck).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anas.
OX NCBI_TaxID=8839;
RN [1]
RP SEQUENCE.
RX TISSUE=Feather;
RA Arai K.M., Takahashi R., Yokote Y., Akahane K.;
RT "The primary structure of feather keratins from duck (Anas
RT platyrhynchos) and pigeon (Columba livia).";
RL Biochim. Biophys. Acta 873:6-12(1986).
CC -!- SUBUNIT: THE AVIAN KERATINS (F-KER, S-KER, C-KER AND B-KER) ARE A
CC COMPLEX MIXTURE OF VERY SIMILAR POLYPEPTIDES.
CC -!- SIMILARITY: BELONGS TO THE AVIAN KERATIN FAMILY.
DR PIR; A24506; KRDKF4.

```

DR InterPro; IPR003461; Keratin.
DR Pfam; PF02422; Keratin; 1.
KW Keratin; Feather; Multigene family; Acetylation.
FT MOD RES 1 1 ACETYLATION.
SQ SEQUENCE 95 AA; 9662 MW; B90FC4CA43805817 CRC64;
Query Match 30.0%; Score 6; DB 1; Length 95;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 13 PSPVVV 18
Db 35 PSPVVV 40

Search completed: November 21, 2003, 20:59:57
Job time : 5.85 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:45:57 ; Search time 24.4 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-80

Perfect score: 20

Sequence: 1 AKSVGPVYCFPSPVVVG 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	206	12 Q81571	hepatitis c
2	20	100.0	778	12 Q41185	hepatitis c
3	20	100.0	778	12 Q41184	hepatitis c
4	20	100.0	2436	12 Q81756	hepatitis c
5	20	100.0	3011	12 Q31F25	hepatitis c
6	20	100.0	3011	12 Q36579	hepatitis c
7	20	100.0	3011	12 Q36610	hepatitis c
8	20	100.0	3011	12 Q34463	hepatitis c
9	20	100.0	3011	12 Q9EL58	hepatitis c
10	20	100.0	3011	12 Q36609	hepatitis c
11	20	100.0	3011	12 Q36608	hepatitis c
12	20	100.0	3032	12 Q99IB4	hepatitis c
13	20	100.0	3033	12 Q31Z22	hepatitis c
14	18	90.0	137	12 Q914Q6	hepatitis c
15	18	90.0	137	12 Q914Q5	hepatitis c
16	18	90.0	137	12 Q914Q8	hepatitis c

17	18	90.0	137	12 Q914Q4	hepatitis c
18	18	90.0	137	12 Q914Q0	hepatitis c
19	18	90.0	137	12 Q914Q3	hepatitis c
20	18	90.0	137	12 Q914Q2	hepatitis c
21	18	90.0	137	12 Q914Q1	hepatitis c
22	18	90.0	137	12 Q914Q7	hepatitis c
23	18	90.0	335	12 Q91HQ5	hepatitis c
24	18	90.0	364	12 Q37937	hepatitis c
25	18	90.0	424	12 Q98UN6	hepatitis c
26	18	90.0	3008	12 Q39929	hepatitis c
27	18	90.0	3013	12 Q92530	hepatitis c
28	18	90.0	3016	12 Q92531	hepatitis c
29	18	90.0	3018	12 Q39927	hepatitis c
30	18	90.0	3019	12 Q92529	hepatitis c
31	18	90.0	3021	12 Q81258	hepatitis c
32	18	90.0	3032	12 Q99IB2	hepatitis c
33	18	90.0	3032	12 Q9QA9	hepatitis c
34	18	90.0	3033	12 Q9DHD6	hepatitis c
35	18	90.0	3033	12 Q9QA8	hepatitis c
36	18	90.0	3033	12 Q99IB8	hepatitis c
37	18	90.0	3033	12 Q91Z45	hepatitis c
38	18	90.0	3033	12 Q99IB6	hepatitis c
39	18	90.0	3033	12 Q9QA7	hepatitis c
40	18	90.0	3033	12 Q91Z43	hepatitis c
41	18	90.0	3033	12 Q91Z44	hepatitis c
42	18	90.0	3033	12 Q9Q9B0	hepatitis c
43	18	90.0	3033	12 Q91Z46	hepatitis c
44	18	90.0	3033	12 Q99IB3	hepatitis c
45	17	85.0	137	12 Q91458	hepatitis c

ALIGNMENTS

RESULT 1

Q81571 ID Q81571 PRELIMINARY; PRT; 206 AA.
AC Q81571;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE NS2 protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Hutchinson;
RX MEDLINE=91013116; PubMed=2170712;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA Yoshizawa H.;
RT "The 5'-terminal sequence of the hepatitis C virus genome.";
RL Jpn. J. Exp. Med. 60:167-177(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Hutchinson;
RA Inchauspe G., Zebedee S.L., Nasoff M.S., Sugitani M., Abe K.,
RA Prince A.M.;
RT "Cloning and nucleotide sequence analysis of structural and
RT nonstructural regions of the hutchinson strain of hepatitis C.";
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Hutchinson;
RX MEDLINE=89222455; PubMed=2496467;
RA Kuo G., Choo Q.-L., Alter H.J., Gitnick G.L., Redeker A.G.,
RA Purcell R.H., Miyamura T., Dienstag J.L., Alter M.J., Stevens C.E.,
RA Testmeier G.E., Bonino F., Colombo M., Lee W.-S., Kuo C., Berger K.,
RA Shuster J.R., Overby L.R., Bradley D.W., Houghton M.;
RT "An assay for circulating antibodies to a major etiologic virus of
RT human non-A, non-B hepatitis.";
RL Science 244:362-365(1992).

```
DR EMBL; M55971; AAA45615.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
FT NON_TER 1
FT NON_TER 206
SQ SEQUENCE 206 AA; 22807 MW; 7FC6174D92325295 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 206;
Best Local Similarity 100.0%; Pred. No. 6.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20
DB 9 AKSVCGPVYCFPTSPVVVGT 28

RESULT 2
Q04185 ID Q04185 PRELIMINARY; PRT; 778 AA.
AC Q04185; Q81810;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Structural (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91195357; PubMed=1849654;
RA Ogata N., Alter H.J., Miller R.H., Purcell R.H.;
RT "Nucleotide sequence and mutation rate of the H strain of hepatitis C
virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:3392-3396 (1991).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; M62381; AAB02127.1; -.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane. STRUCTURAL PROTEIN.
FT CHAIN 514 >778
FT NON_TER 778
SQ SEQUENCE 778 AA; 85143 MW; F7709172CD03E39B CRC64;

Query Match 100.0%; Score 20; DB 12; Length 778;
Best Local Similarity 100.0%; Pred. No. 1.9e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20
DB 499 AKSVCGPVYCFPTSPVVVGT 518

RESULT 4
Q81756 ID Q81756 PRELIMINARY; PRT; 2436 AA.
AC Q81756;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Choo Q.-L., Richman K., Han J.;
RT "The nucleotide sequence of the Hepatitis C viral genome.";
RL Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.
DR EMBL; M32084; AAA45677.1; -.
DR HSP; P27958; IAI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.

Query Match 100.0%; Score 20; DB 12; Length 778;
Best Local Similarity 100.0%; Pred. No. 1.9e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20
DB 499 AKSVCGPVYCFPTSPVVVGT 518

RESULT 3
Q04184 ID Q04184 PRELIMINARY; PRT; 778 AA.
AC Q04184; Q81810;
```

DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDc; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
DR KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA, 327124 MW, 2489CE74AC864E58 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 5.7e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFTSPVVVGVT 20
|||||
DB 499 AKSVCGPVYCFTSPVVVGVT 518

RESULT 6
O36579 PRELIMINARY; PRT; 3011 AA.

ID AC O36579;
AC O36579;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DI 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97373636; PubMed=9228008;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstein S.M., Rice C.M.;
RT "Transmission of hepatitis C by intrahepatic inoculation with transcribed RNA.";
RL Science 277:570-574 (1997).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AF009606; AAC66324.1; -.
DR HSSP; P27958; 1HEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.

```
DR Pfam: PF01538; HCV NS2; 1.
DR Pfam: PF02907; HCV NS3; 1.
DR Pfam: PF01006; HCV NS4a; 1.
DR Pfam: PF01001; HCV NS4b; 1.
DR Pfam: PF01506; HCV NS5a; 1.
DR Pfam: PF00271; helicase C; 1.
DR Pfam: PF00998; Viral RdRp; 1.
DR Pfam: PF00271; helicase C; 1.
DR Pfam: PF00998; Viral RdRp; 1.
DR ProDom: PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydroxylase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327182 MW; E2E0EE809C63C1B9 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 5.7e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20
DB 499 AKSVCGPVYCFPTSPVVVGT 518

RESULT 7
ID O36610 PRELIMINARY; PRT; 3011 AA.
AC O36610;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-WAR-2003 (TREMBlrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=63746;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97385173; PubMed=9238047;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus
RT are infectious when directly transfected into the liver of a
RT chimpanzee."
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF011753; AB67038.1; -.
DR HSP; P27958; IHEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002531; HCV NS2.
DR InterPro; IPR002531; HCV NS3.
DR InterPro; IPR004109; HCV NS4.
DR InterPro; IPR000745; HCV NS4b.
DR InterPro; IPR001490; HCV NS4a.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
```

```
DR Pfam: PF01001; HCV NS4b; 1.
DR Pfam: PF01506; HCV NS5a; 1.
DR Pfam: PF00271; helicase C; 1.
DR Pfam: PF00998; Viral RdRp; 1.
DR ProDom: PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydroxylase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327220 MW; 293F91430A0D4067 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 5.7e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20
DB 499 AKSVCGPVYCFPTSPVVVGT 518

RESULT 8
ID Q03463 PRELIMINARY; PRT; 3011 AA.
AC Q03463;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=91013116; PubMed=2170712;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA Yoshizawa H.;
RT "The 5'-terminal sequence of the hepatitis C virus genome."
RL Jpn. J. Exp. Med. 60:167-177(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=92044440; PubMed=1658196;
RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
RA Miyakawa Y., Mayumi M.;
RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions."
RL J. Gen. Virol. 72:2697-2704(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=93117120; PubMed=1335573;
RA Okamoto H., Kanai N., Mishihiro S.;
RT "Full-length nucleotide sequence of a Japanese hepatitis C virus
RT isolate (HC-J1) with high homology to USA isolates."
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RA Okamoto H.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=94174722; PubMed=7510436;
RA Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,
RA Inchausti G.;
RT "Characterization and mapping of a B-cell immunogenic domain in
RT hepatitis C virus E2 glycoprotein using a yeast peptide library.";
```



```

DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA pol DS ps.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007094; RNA pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327261 MW; 10D1C9702CA9B5DC CRC64;

Query Match 100.0%; Score 20; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 5.7e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20
DB 499 AKSVCGPVYCFPTSPVVVGT 518

RESULT 11
O36608 PRELIMINARY; PRT; 3011 AA.
AC O36608;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus strain H77.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=63746;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97385173; PubMed=9238047;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus
RT are infectious when directly transfected into the liver of a
RT chimpanzee."
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF011751; AAB67036.1; -.
DR HSP; P27958; IHEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.

```

```

DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA pol_DS_Ps.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327112 MW; 0B75E6B81CB5C198 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 5.7e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKSVCGPVYCFPTSPVVVGT 20
DB 499 AKSVCGPVYCFPTSPVVVGT 518

RESULT 12
Q991B4 PRELIMINARY; PRT; 3032 AA.
AC Q991B4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=21316767; PubMed=11424123;
RA Kato T., Furusaka A., Miyamoto M., Date T., Yasui K., Hiramoto J.,
RA Nagayama K., Tanaka T., Wakita T.;
RT "Sequence analysis of hepatitis C virus isolated from a fulminant
RT hepatitis patient."
RL J. Med. Virol. 64:334-339(2001).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AB047643; BAB32876.1; -.
DR HSP; P27958; IHEI.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV env.
DR InterPro; IPR002519; HCV NS1.
DR InterPro; IPR002531; HCV NS2.

```

DR	InterPro: IPR007094; RNA_pol_PSVir.
DR	Pfam: PF01543; HCV_capsid; 1.
DR	Pfam: PF01542; HCV_core; 1.
DR	Pfam: PF01539; HCV_env; 1.
DR	Pfam: PF01560; HCV_NS1; 1.
DR	Pfam: PF01538; HCV_NS2; 1.
DR	Pfam: PF02307; HCV_NS3; 1.
DR	Pfam: PF01006; HCV_NS4a; 1.
DR	Pfam: PF01001; HCV_NS4b; 1.
DR	Pfam: PF01506; HCV_NS5a; 1.
DR	Pfam: PF00998; Viral_RdRP; 1.
DR	ProDom: PD186062; HCV_NS1; 1.
DR	SMART; SM00487; DEXDC; 1.
DR	PROSITE; PS00190; CYTOCHROME C; 1.
DR	PROSITE; PS05057; RDRP_POSITIVE; 1.
DR	PROSITE; PS05021; RDRP_VIRAL; 1.
KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW	Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ	SEQUENCE 3033 AA; 329531 MW; BE13557FABA1E87 CRC64;
Query Match 100.0%; Score 20; DB 12; Length 3033;	
Best Local Similarity 100.0%; Pred. No. 5.7e-12;	
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 AKSVCGPVYCFTPSPVVVGT 20
DB	501 AKSVCGPVYCFTPSPVVVGT 520
RESULT 14	
Q91406	PRELIMINARY; PRT; 137 AA.
ID	Q91406
AC	Q91406;
DT	01-DEC-2001 (TrEMBLrel. 19, Created)
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE	Genome polyprotein (Fragment).
OS	Hepatitis C virus.
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC	Hepacivirus.
OX	NCBI_TaxID=11103;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=wt;
RT	Lyra A.C., Fan X., Di Bisceglie A.M.;
RT	"Genetic diversity of two putative cdb1 binding regions of HCV before
RT	and after liver transplantation.";
RL	Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF431882; AAL30731.1; -
DR	InterPro; IPR002531; HCV_NS1.
DR	Pfam; PF01560; HCV_NS1; 1.
DR	ProDom; PD186062; HCV_NS1; 1.
KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW	Polyprotein; Transmembrane.
FT	NON_TER 1
FT	NON_TER 137
SQ	SEQUENCE 137 AA; 15042 MW; 08E5232119C91A09 CRC64;
Query Match 90.0%; Score 18; DB 12; Length 137;	
Best Local Similarity 100.0%; Pred. No. 5.3e-11;	
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	3 SVCGPVCFTPSPVVVGT 20
DB	23 SVCGPVCFTPSPVVVGT 40
RESULT 15	
Q91405	PRELIMINARY; PRT; 137 AA.
ID	Q91405
AC	Q91405;
DT	01-DEC-2001 (TrEMBLrel. 19, Created)
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RC STRAIN=wt;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
RT and after liver transplantation."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431883; AAL30732.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 15070 MW; 08F2132F29D01A09 CRC64;
Query Match 90.0%; Score 18; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 5.3e-11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3 SVCGPVYCFPTSPVVVGT 20
Db |||||
23 SVCGPVYCFPTSPVVVGT 40

Search completed: November 21, 2003, 21:08:18
Job time : 25.45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:22 ; Search time 32.1 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-81

Perfect score: 20
Sequence: 1 PSPVVVGTDRSGAPTYSWG 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

1:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
24:	/SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	16	AA84510
2	20	100.0	20	17	AA84510
3	20	100.0	20	23	AA018710
4	20	100.0	20	23	AA018710
5	20	100.0	30	23	AAU84632
6	20	100.0	31	23	AAU84632
7	20	100.0	36	24	ABP70588
8	20	100.0	37	23	ABP70588
9	20	100.0	43	23	ABP70588

10	20	100.0	96	23	ABP77258
11	20	100.0	176	23	ABP77257
12	20	100.0	179	21	AA018529
13	20	100.0	250	23	AAU79220
14	20	100.0	251	24	ABP96704
15	20	100.0	254	22	AA088043
16	20	100.0	271	22	AA088045
17	20	100.0	271	22	AA088046
18	20	100.0	271	22	AA088046
19	20	100.0	278	24	ABP96703
20	20	100.0	278	24	ABP96705
21	20	100.0	305	18	AAW00929
22	20	100.0	333	14	AA040118
23	20	100.0	337	16	AA040117
24	20	100.0	350	23	AA018991
25	20	100.0	363	22	AA088042
26	20	100.0	363	24	AA032880
27	20	100.0	363	24	ABP55567
28	20	100.0	367	14	AA040115
29	20	100.0	367	16	AA040118
30	20	100.0	377	16	AA040118
31	20	100.0	397	16	AA040117
32	20	100.0	399	14	AA040117
33	20	100.0	402	14	AA034438
34	20	100.0	402	14	AA034439
35	20	100.0	409	14	AA033995
36	20	100.0	410	16	AA079227
37	20	100.0	417	16	AA079228
38	20	100.0	434	16	AA079219
39	20	100.0	441	16	AA079230
40	20	100.0	447	16	AA079229
41	20	100.0	453	16	AA079225
42	20	100.0	454	10	AA090183
43	20	100.0	454	10	AA092049
44	20	100.0	454	21	AA018526
45	20	100.0	463	14	AA033588
	20	100.0	463	14	AA033588

ALIGNMENTS

RESULT 1
AA84510
ID AA84510 standard; peptide; 20 AA.
AC AA84510;
XX
XX
DT 06-JAN-1997 (first entry)
XX
XX
DE Hepatitis C virus peptide NS1-11 (residues 511-530).
XX
XX
KW Hepatitis C virus; HCV; immunogen; non-structural region; NS1;
KW immunodominant; T cell epitope; vaccine.
XX
XX
OS Hepatitis C virus.
XX
XX
PN W09512677-A2.
XX
XX
PD 11-MAY-1995.
XX
XX
PF 28-OCT-1994; 94WO-EP03555.
XX
XX
PR 04-NOV-1993; 93EP-0402718.
XX
XX
PA (INNO-) INNOGENETICS NV.
XX
XX
PI Deleys R, Leroux-Roels G, Maertens G;
XX
XX
DR WPI; 1995-193822/25.
XX
XX
PT Hepatitis C Virus immunogenic polypeptide contg. a T-cell
PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in
PT production of vaccines, therapeutic agents, etc.

```
XX Example 4; Page 51; 105pp; English.
PS
CC A series of overlapping peptides (including the present sequence) was
CC synthesised based on sequences in the core, E1 and E2/NS1 regions of
CC hepatitis C virus. The peptides were used as antigens in lympho-
CC proliferative assays to identify the main T-cell epitopes.
XX
SQ Sequence 20 AA;
Query Match 100.0%; Score 20; DB 16; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PSPVVVGTTDRSGAPTYSWG 20
DB 1 PSPVVVGTTDRSGAPTYSWG 20
RESULT 2
AAR91004
ID AAR91004 standard; peptide; 20 AA.
AC AAR91004;
XX
DT 25-SEP-1996 (first entry)
XX
DE HCV E2 peptide E2-11B for competition studies.
XX
KW HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;
KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.
XX
OS Synthetic.
XX
PN WO9604385-A2.
XX
PD 15-FEB-1996.
XX
PF 31-JUL-1995; 95WO-EP03031.
XX
PR 29-JUL-1994; 94EP-0870132.
XX
PA (INNO-) INNOGENETICS NV.
XX
PI Bosman F, Buyse M, De Martynoff G, Maertens G;
XX
DR WPI; 1996-129401/13.
XX
PT Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope
PT proteins - in presence of disulphide bond cleavage agent, to
PT produce proteins suitable for direct use in vaccines or diagnostic
PT assays of HCV
XX
XX Example 7; Page 67; 146pp; English.
XX
CC AAR90965-R30980 and AAR90995-R31015 represent synthetic hepatitis C
CC virus (HCV) E1 and E2 peptides used in competition studies. This
CC sequence represents a synthetic E2 peptide, and corresponds to residues
CC 511-530 of the E2 protein sequence. These sequences are useful for in
CC vitro monitoring of HCV disease, or prognosis of the response to
CC interferon treatment of patients suffering from HCV infection. These
CC sequences compete with the proteins produced by AAT12704-T12709 and
CC AAT12963-T12974, which are included in vectors for the production of
CC recombinant E1, E2, and E1/E2. The recombinant proteins can then be
CC isolated and purified by carrying out a disulphide bond cleavage, or a
CC reduction step with a disulphide bond cleavage agent, after lysis of
CC recombinant host cells. The constructs containing the purified HCV
CC envelope proteins can be used for vaccinating humans against HCV, for in
CC vitro detection of HCV antibodies in a sample, and in a serotyping assay
CC for detecting one or more serological types of HCV present in a
CC biological sample. The constructs can also be immobilised on a solid
CC substrate and incorporated into a reversed phase hybridisation assay for
CC determining the presence or the genotype of HCV. The new purification
```

```
CC method preserves the conformation of the recombinantly expressed E1, E2
CC and E1/E2, and eliminates contaminating proteins. Antigens isolated
CC using this method are more reactive with human sera than those isolated
CC by known techniques.
XX
SQ Sequence 20 AA;
Query Match 100.0%; Score 20; DB 17; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PSPVVVGTTDRSGAPTYSWG 20
DB 1 PSPVVVGTTDRSGAPTYSWG 20
RESULT 3
AAO18710
ID AAO18710 standard; Peptide; 20 AA.
XX
AC AAO18710;
XX
DT 24-OCT-2002 (first entry)
XX
DE Hepatitis C virus E2 protein derived peptide E2-11B.
XX
KW Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide;
KW immunostimulant; vaccine.
XX
OS Hepatitis C virus.
XX
PN WO200255548-A2.
XX
PD 18-JUL-2002.
XX
PF 11-JAN-2002; 2002WO-EP00219.
XX
PR 11-JAN-2001; 2001US-260699P.
PR 30-AUG-2001; 2001US-315768P.
XX
PA (INNO-) INNOGENETICS NV.
XX
PI Maertens G, Bosman F, Buyse M;
XX
DR WPI; 2002-599657/64.
XX
PT New therapeutic vaccine compositions comprising at least one purified
PT recombinant hepatitis C virus (HCV) single or specific oligomeric
PT recombinant envelope protein E1 or E2, useful for immunizing humans
PT from HCV infection.
XX
PS Example 7; Page 228; 243pp; English.
XX
XX The present invention relates to new therapeutic vaccine compositions for
XX inducing hepatitis C virus (HCV)-specific antibodies, comprising a
XX composition containing at least one purified recombinant HCV single or
XX specific oligomeric recombinant envelope proteins selected from an E1 and
XX an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are
XX useful for inducing HCV-specific antibodies or for immunising humans
XX against HCV. The recombinant HCV E1 and/or E2 proteins are useful as
XX vaccines or therapeutics, in HCV screening and confirmatory antibody
XX tests, for raising antibodies, in the preparation of medicament, and for
XX in vitro monitoring of HCV disease or prognosing the response to
XX treatment of patients suffering from HCV infection. The present sequence
XX is a peptide derived from the proteins of the invention.
SQ Sequence 20 AA;
Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PSPVVVGTTDRSGAPTYSWG 20
```

```

Db      1 PSPVVVGTTDRSGAPTSWMG 20
|||||
RESULT 4
ABB77220
ID  ABB77220 standard; Protein; 28 AA.
XX
XX  ABB77220;
AC
XX  28-JUN-2002 (first entry)
DT
XX
DE  Selected interacting domain (SID) polypeptide 6.
XX
XX  SID; selected interacting domain; HCV; hepatitis C virus;
KW  liver disease; liver cancer; virucide; hepatotropic; antiinflammatory;
KW  antibacterial.
XX
XX  Hepatitis C virus strain H77.
OS
XX  EP1178116-A1.
PN
XX
XX  06-FEB-2002.
PD
XX
XX  03-AUG-2000; 2000EP-0402225.
PF
XX
XX  03-AUG-2000; 2000EP-0402225.
PR
XX
XX  (HYBR-) HYBRIGENICS SA.
PA
XX  Legrain P, Whiteside S, Wojcik J;
PI
XX  WPI; 2002-208115/27.
DR
XX  N-PSDB; ABL55552.
XX
XX  New selected interacting domain polypeptides and polynucleotides,
PT  useful for treating or preventing infections or pathologies caused by
PT  hepatitis C virus (HCV) or those linked to HCV infection -
XX
XX  Claim 7; SEQ ID 6; 61pp + sequence listing; English.
XX
XX  The invention relates to nucleic acids encoding polypeptides which are
CC  termed SID polypeptides (selected interacting domain). These polypeptides
CC  are the final products of a double selection method involving a first
CC  step of selection of Hepatitis C virus (HCV)-derived polynucleotides
CC  through a two-hybrid system, and a second selection step involving an
CC  alignment between the different polynucleotides selected at the first
CC  step. The activity of polypeptides of the invention may be described as,
CC  virucide, hepatotropic, antiinflammatory and antibacterial. The
CC  polypeptide, polynucleotide and compositions comprising them are useful
CC  for treating or preventing viral or a bacterial infection, specifically
CC  infections or pathologies caused by HCV, or those pathologies linked to
CC  HCV infection. These may include liver disease and liver cancer. The
CC  current sequence represents a selected interacting domain (SID)
CC  polypeptide.
CC  Note: The sequence data for this patent is not represented in the
CC  specification, but is based on sequence information supplied by the
CC  European Patent Office.
XX
XX  Sequence 28 AA;
SQ
Query Match      100.0%; Score 20; DB 23; Length 28;
Best Local Similarity 100.0%; Pred. No. 2.9e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 PSPVVVGTTDRSGAPTSWMG 20
|||||
Db      1 PSPVVVGTTDRSGAPTSWMG 20
|||||
RESULT 5
AAU84632
ID  AAU84632 standard; Peptide; 30 AA.
XX
XX  08-MAY-2002 (first entry)
DT
XX
XX  HCV HepC1a segment 35.
DE
XX
XX  Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
KW  viral infection; human immunodeficiency virus; melanoma;
KW  bacterial infection; Salmonella; Legionella; parasitic infection;
KW  Trypanosoma; Toxoplasma; Giardia.
XX
XX  Hepatitis C virus.
OS
XX  WO200190197-A1.
PN
XX
XX  29-NOV-2001.
PD
XX
XX  25-MAY-2001; 2001WO-AU00622.
PF
XX
XX  26-MAY-2000; 2000AU-0007761.
PR
XX
XX  (AUSU ) UNIV AUSTRALIAN NAT.
PA
XX  Thomson SA, Ramshaw IA;
PI
XX
XX  WPI; 2002-147575/19.
DR
XX  N-PSDB; ABK36470.
XX
XX  New synthetic polypeptides having several different segments of at
PT  least one parent polypeptide linked together differently compared to
PT  the linkage in the parent polypeptide, for inducing immune response
PT  against a pathogen or cancer -
XX
XX  Example 2; Fig 26; 364pp; English.
XX
XX  The invention relates to a new synthetic polypeptide (I) comprising
CC  several different segments of at least one parent polypeptide linked
CC  together in a different relationship relative to their linkage in the
CC  parent polypeptide to impede, abrogate or otherwise alter at least one
CC  function associated with the parent polypeptide and for inducing an
CC  immune response against a pathogen or cancer. Also included are a
CC  synthetic polynucleotide encoding and a computer system for
CC  designing the synthetic polypeptides. The synthetic polypeptides and
CC  polynucleotides are referred to as a Savine. The synthetic polypeptide is
CC  useful for modulating immune responses preferably directed against a
CC  pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix,
CC  colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone
CC  liver, oesophagus, brain, testicle, uterus), as potentiating agents.
CC  Compositions comprising the polypeptide may be used in the treatment or
CC  prophylaxis against viral (such as infections caused by HIV (human
CC  immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
CC  virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
CC  (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,
CC  salmonella, Streptococcal, Legionella and Mycobacterium or parasitic
CC  (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
CC  Trypanosoma, Toxoplasma and Giardia) infections. The present
CC  sequence is a peptide derived from a parent protein used to
CC  construct a savine of the invention.
XX
XX  Sequence 30 AA;
SQ
Query Match      100.0%; Score 20; DB 23; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 PSPVVVGTTDRSGAPTSWMG 20
|||||
Db      3 PSPVVVGTTDRSGAPTSWMG 22
|||||
RESULT 6
ABB77229
```

ID ABB77229 standard; Protein; 31 AA.
XX
AC ABB77229;
XX
DT 28-JUN-2002 (first entry)
XX
DE Selected interacting domain (SID) polypeptide 15.
XX
KW SID; selected interacting domain; HCV; hepatitis C virus;
KW liver disease; liver cancer; virucide; hepatotropic; antiinflammatory;
KW antibacterial.
XX
OS Hepatitis C virus strain H77.
XX
FN EP1178116-A1.
XX
PD 06-FEB-2002.
XX
PF 03-AUG-2000; 2000EP-0402225.
XX
PR 03-AUG-2000; 2000EP-0402225.
XX
PA (HYBR-) HYBRIGENICS SA.
XX
PI Legrain P, Whiteside S, Wojcik J;
XX
DR WPI; 2002-208115/27.
DR N-PSDB; ABL55561.
XX
XX New selected interacting domain polypeptides and polynucleotides,
PT useful for treating or preventing infections or pathologies caused by
PT hepatitis C virus (HCV) or those linked to HCV infection -
XX
PS Claim 7; SEQ ID 15; 61pp + sequence listing; English.
XX
CC The invention relates to nucleic acids encoding polypeptides which are
CC termed SID polypeptides (selected interacting domain). These polypeptides
CC are the final products of a double selection method involving a first
CC step of selection of Hepatitis C virus (HCV)-derived polynucleotides
CC through a two-hybrid system, and a second selection step involving an
CC alignment between the different polynucleotides selected at the first
CC step. The activity of polypeptides of the invention may be described as,
CC virucide, hepatotropic, antiinflammatory and antibacterial. The
CC polypeptide, polynucleotide and compositions comprising them are useful
CC for treating or preventing viral or a bacterial infection, specifically
CC infections or pathologies caused by HCV, or those pathologies linked to
CC HCV infection. These may include liver disease and liver cancer. The
CC current sequence represents a selected interacting domain (SID)
CC polypeptide.
CC Note: The sequence data for this patent is not represented in the
CC specification, but is based on sequence information supplied by the
CC European Patent Office.
XX
SQ Sequence 31 AA;
Query Match 100.0%; Score 20; DB 23; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 PSPVVVGTTDRSGAPTYSWG 20
Db 1 PSPVVVGTTDRSGAPTYSWG 20
RESULT 7
ABP70588
ID ABB70588 standard; peptide; 36 AA.
XX
AC ABB70588;
XX
DT 22-APR-2003 (first entry)
XX
DE Hepatitis C virus E2 glycoprotein derived antigenic peptide.

XX
KW Flavivirus; vector; envelope protein; peptide delivery; immune response;
KW tumour-associated antigen; immunotherapy; cancer; HCV; E2 glycoprotein.
XX
OS Hepatitis C virus.
XX
FN WO2002102828-A2.
XX
PD 27-DEC-2002.
XX
PF 31-MAY-2002; 2002WO-US17374.
XX
PR 01-JUN-2001; 2001US-295265P.
XX
PA (ACAM-) ACAMBIS INC.
XX
FI Kleanthous H, Oros L, Miller C;
XX
DR WPI; 2003-167480/16.
XX
PT Novel flavivirus vector useful for delivering foreign peptides,
PT comprises an envelope protein that comprises a foreign peptide -
XX
PS Example; Page 18; 34pp; English.
XX
CC The specification describes a flavivirus vector, comprising an envelope
CC protein that comprises a foreign peptide. The vector is useful for
CC delivery of a peptide, e.g. an antigen to a patient. It can be
CC administered to induce an immune response to a pathogen or tumour from
CC which the antigen is derived. It also useful for delivering
CC tumour-associated antigens for use in immunotherapeutic methods against
CC cancer. ABP70571-96 represent B and T cell epitopes from viruses.
CC They may be expressed using the vectors of the invention.
XX
SQ Sequence 36 AA;
Query Match 100.0%; Score 20; DB 24; Length 36;
Best Local Similarity 100.0%; Pred. No. 3.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 PSPVVVGTTDRSGAPTYSWG 20
Db 1 PSPVVVGTTDRSGAPTYSWG 20
RESULT 8
ABB77218
ID ABB77218 standard; Protein; 37 AA.
XX
AC ABB77218;
XX
DT 28-JUN-2002 (first entry)
XX
DE Selected interacting domain (SID) polypeptide 4.
XX
KW SID; selected interacting domain; HCV; hepatitis C virus;
KW liver disease; liver cancer; virucide; hepatotropic; antiinflammatory;
KW antibacterial.
XX
OS Hepatitis C virus strain H77.
XX
FN EP1178116-A1.
XX
PD 06-FEB-2002.
XX
PF 03-AUG-2000; 2000EP-0402225.
XX
PR 03-AUG-2000; 2000EP-0402225.
XX
PA (HYBR-) HYBRIGENICS SA.
XX
PI Legrain P, Whiteside S, Wojcik J;
XX

DR WPI; 2002-208115/27.
 DR N-PSDB; ABL55550.
 XX
 XX New selected interacting domain polypeptides and polynucleotides.
 PT useful for treating or preventing infections or pathologies caused by
 PT hepatitis C virus (HCV) or those linked to HCV infection -
 XX
 PS Claim 7; SEQ ID 4; 61pp + sequence listing; English.
 XX
 XX The invention relates to nucleic acids encoding polypeptides which are
 CC termed SID polypeptides (selected interacting domain). These polypeptides
 CC are the final products of a double selection method involving a first
 CC step of selection of Hepatitis C virus (HCV)-derived polynucleotides
 CC through a two-hybrid system, and a second selection step involving an
 CC alignment between the different polynucleotides selected at the first
 CC step. The activity of polypeptides of the invention may be described as,
 CC virucide, hepatotropic, antiinflammatory and antibacterial. The
 CC polypeptide, polynucleotide and compositions comprising them are useful
 CC for treating or preventing viral or a bacterial infection, specifically
 CC infections or pathologies caused by HCV, or those pathologies linked to
 CC HCV infection. These may include liver disease and liver cancer. The
 CC current sequence represents a selected interacting domain (SID)
 CC polypeptide.
 CC Note: The sequence data for this patent is not represented in the
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.
 XX
 SQ Sequence 37 AA;
 Query Match 100.0%; Score 20; DB 23; Length 37;
 Best Local Similarity 100.0%; Pred. No. 3.7e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PSPVVVGTTDRSGAPTYSWG 20
 DB 1 PSPVVVGTTDRSGAPTYSWG 20
 RESULT 9
 ABB77236
 ID ABB77236 standard; Protein; 43 AA.
 XX
 AC ABB77236;
 XX
 DT 28-JUN-2002 (first entry)
 XX
 DE Selected interacting domain (SID) polypeptide 22.
 XX
 KW SID; selected interacting domain; HCV; hepatitis C virus;
 KW liver disease; liver cancer; virucide; hepatotropic; antiinflammatory;
 KW antibacterial.
 XX
 OS Hepatitis C virus strain H77.
 XX
 PN EP1178116-A1.
 XX
 PD 06-FEB-2002.
 XX
 PF 03-AUG-2000; 2000EP-0402225.
 XX
 PR 03-AUG-2000; 2000EP-0402225.
 XX
 PA (HYBR-) HYBRIGENICS SA.
 XX
 PI Legrain P, Whiteside S, Wojcik J;
 XX
 DR WPI; 2002-208115/27.
 DR N-PSDB; ABL55568.
 XX
 XX New selected interacting domain polypeptides and polynucleotides,
 PT useful for treating or preventing infections or pathologies caused by
 PT hepatitis C virus (HCV) or those linked to HCV infection -
 XX

PS Claim 7; SEQ ID 22; 61pp + sequence listing; English.
 XX
 XX The invention relates to nucleic acids encoding polypeptides which are
 CC termed SID polypeptides (selected interacting domain). These polypeptides
 CC are the final products of a double selection method involving a first
 CC step of selection of Hepatitis C virus (HCV)-derived polynucleotides
 CC through a two-hybrid system, and a second selection step involving an
 CC alignment between the different polynucleotides selected at the first
 CC step. The activity of polypeptides of the invention may be described as,
 CC virucide, hepatotropic, antiinflammatory and antibacterial. The
 CC polypeptide, polynucleotide and compositions comprising them are useful
 CC for treating or preventing viral or a bacterial infection, specifically
 CC infections or pathologies caused by HCV, or those pathologies linked to
 CC HCV infection. These may include liver disease and liver cancer. The
 CC current sequence represents a selected interacting domain (SID)
 CC polypeptide.
 CC Note: The sequence data for this patent is not represented in the
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.
 XX
 SQ Sequence 43 AA;
 Query Match 100.0%; Score 20; DB 23; Length 43;
 Best Local Similarity 100.0%; Pred. No. 4.3e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PSPVVVGTTDRSGAPTYSWG 20
 DB 17 PSPVVVGTTDRSGAPTYSWG 36
 RESULT 10
 ABB77258
 ID ABB77258 standard; Protein; 96 AA.
 XX
 AC ABB77258;
 XX
 DT 28-JUN-2002 (first entry)
 XX
 DE HCV bait polypeptide 6.
 XX
 KW SID; selected interacting domain; HCV; hepatitis C virus;
 KW liver disease; liver cancer; virucide; hepatotropic; antiinflammatory;
 KW antibacterial.
 XX
 OS Hepatitis C virus strain H77.
 XX
 PN EP1178116-A1.
 XX
 PD 06-FEB-2002.
 XX
 PF 03-AUG-2000; 2000EP-0402225.
 XX
 PR 03-AUG-2000; 2000EP-0402225.
 XX
 PA (HYBR-) HYBRIGENICS SA.
 XX
 PI Legrain P, Whiteside S, Wojcik J;
 XX
 DR WPI; 2002-208115/27.
 DR N-PSDB; ABL55590.
 XX
 XX New selected interacting domain polypeptides and polynucleotides,
 PT useful for treating or preventing infections or pathologies caused by
 PT hepatitis C virus (HCV) or those linked to HCV infection -
 XX
 PS Claim 26; SEQ ID 82; 61pp + sequence listing; English.
 XX
 XX The invention relates to nucleic acids encoding polypeptides which are
 CC termed SID polypeptides (selected interacting domain). These polypeptides
 CC are the final products of a double selection method involving a first
 CC step of selection of Hepatitis C virus (HCV)-derived polynucleotides
 CC through a two-hybrid system, and a second selection step involving an

CC alignment between the different polynucleotides selected at the first
 CC step. The activity of polypeptides of the invention may be described as,
 CC virucide, hepatotropic, antiinflammatory and antibacterial. The
 CC polypeptide, polynucleotide and compositions comprising them are useful
 CC for treating or preventing viral or a bacterial infection, specifically
 CC infections or pathologies caused by HCV, or those pathologies linked to
 CC HCV infection. These may include liver disease and liver cancer. The
 CC current sequence represents a HCV bait polypeptide.
 CC Note: The sequence data for this patent is not represented in the
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.

SQ Sequence 96 AA;
 Query Match 100.0%; Score 20; DB 23; Length 96;
 Best Local Similarity 100.0%; Pred. No. 8.8e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDRSGAPTYSWG 20
 |||||
 Db 25 PSPVVVGTTDRSGAPTYSWG 44
 |||||

RESULT 11
 ABB77257
 ID ABB77257 standard; Protein; 176 AA.

XX AC ABB77257;

XX DT 28-JUN-2002 (first entry)

XX DE HCV bait polypeptide 5.

XX KW SID; selected interacting domain; HCV; hepatitis C virus;
 XX liver disease; liver cancer; virucide; hepatotropic; antiinflammatory;
 XX antibacterial.

XX OS Hepatitis C virus strain H77.

XX PN EP178116-A1.

XX PD 06-FEB-2002.

XX PF 03-AUG-2000; 2000EP-0402225.

XX PR 03-AUG-2000; 2000EP-0402225.

XX PA (HYBR-) HYBRIGENICS SA.

XX PI Legrain P, Whiteside S, Wojcik J;

XX WPI; 2002-208115/27.

DR N-PSDB; ABL55589.

XX New selected interacting domain polypeptides and polynucleotides,
 PT useful for treating or preventing infections or pathologies caused by
 PT hepatitis C virus (HCV) or those linked to HCV infection -

XX Claim 26; SEQ ID 81; 61pp + sequence listing; English.

XX The invention relates to nucleic acids encoding polypeptides which are
 CC termed SID polypeptides (selected interacting domain). These polypeptides
 CC are the final products of a double selection method involving a first
 CC step of selection of Hepatitis C virus (HCV)-derived polynucleotides
 CC through a two-hybrid system, and a second selection step involving an
 CC alignment between the different polynucleotides selected at the first
 CC step. The activity of polypeptides of the invention may be described as,
 CC virucide, hepatotropic, antiinflammatory and antibacterial. The
 CC polypeptide, polynucleotide and compositions comprising them are useful
 CC for treating or preventing viral or a bacterial infection, specifically
 CC infections or pathologies caused by HCV, or those pathologies linked to
 CC HCV infection. These may include liver disease and liver cancer. The
 CC current sequence represents a HCV bait polypeptide.

CC Note: The sequence data for this patent is not represented in the
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.

SQ Sequence 176 AA;

Query Match 100.0%; Score 20; DB 23; Length 176;
 Best Local Similarity 100.0%; Pred. No. 1.5e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDRSGAPTYSWG 20
 |||||
 Db 43 PSPVVVGTTDRSGAPTYSWG 62
 |||||

RESULT 12
 AAB18529
 ID AAB18529 standard; Protein; 179 AA.

XX AC AAB18529;

XX DT 15-JAN-2001 (first entry)

XX DE Protein encoded by a novel hepatitis C virus cDNA clone 13i.

XX KW Hepatitis C virus; HCV; antisense polynucleotide; polypeptide;
 XX viral infectivity; viral replication.

XX OS Hepatitis C virus.

XX PN EP1034785-A2. 1

XX PD 13-SEP-2000.

XX PF 16-MAR-1990; 2000EP-0109602.

XX PR 17-MAR-1989; 89US-0325338.

XX PR 20-APR-1989; 89US-0341334.

XX PR 18-MAY-1989; 89US-0355002.

XX PR 16-MAR-1990; 90EP-0302866.

XX PA (CHIR) CHIRON CORP.

XX PI Houghton M, Choo Q, Kuo G;

XX WPI; 2000-566891/53.

DR N-PSDB; AAA75285.

XX Novel composition comprising a hepatitis C virus antisense
 PT polynucleotide which is complementary to or corresponds to a sense
 PT strand of the virus genome, and selectively hybridises to it -

XX Example; Fig 5; 75pp; English.

XX The specification describes a pharmaceutical composition which
 CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
 CC HCV is characterized by a positive stranded RNA genome which has
 CC 40% homology at the polypeptide level to a HCV polypeptide. The
 CC antisense polynucleotide binds to cellular polynucleotides which
 CC enhance and/or are required for viral infectivity, replicative
 CC ability or chronicity. The antisense polynucleotides may also be
 CC designed to bind with high specificity, to be of increased stability,
 CC to be stable and to have low toxicity. The composition also comprises
 CC an agent which causes viral RNA to be inactive. The composition
 CC is used for preventing HCV replication in a system. The present
 CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
 CC course of the invention.

XX SQ Sequence 179 AA;

Query Match 100.0%; Score 20; DB 21; Length 179;
 Best Local Similarity 100.0%; Pred. No. 1.5e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTDRSGAPTYSWG 20
 |||||
 DB 1 PSPVVVGTDRSGAPTYSWG 20

RESULT 13
 AAU79220
 ID AAU79220 standard; Protein; 250 AA.
 XX
 AC AAU79220;
 XX
 DT 15-JUL-2002 (first entry)
 XX
 DE Hepatitis C Virus (HCV) delta-delta E2 genotype 1a protein.
 XX
 KW Hepatitis C Virus E2; HCV E2; virucide; hepatotropic; IL-2;
 KW antinflammatory; HCV infection; interleukin-2; gamma-interferon;
 KW granulocyte macrophage-colony stimulating factor; GM-CSF;
 KW delta-delta E2 genotype 1a.
 XX
 OS Hepatitis C Virus.
 XX
 PN WO200222155-A1.
 XX
 PD 21-MAR-2002.
 XX
 PF 13-SEP-2001; 2001WO-US28767.
 XX
 PR 13-SEP-2000; 2000US-230927P.
 XX
 PA (HAWAII) HAWAII BIOTECHNOLOGY GROUP INC.
 XX
 PI Nakano ET, Clements DE, Humphreys T;
 XX
 DR WPI; 2002-383102/41.
 DR N-PSDB; ABK49386.
 XX
 PT New immunogenic polypeptide comprising hepatitis C virus E2 polypeptide
 PT useful for treating hepatitis C virus infection and for providing
 PT immune protection against virus infection -
 XX
 PS Claim 4; Fig 6; 84pp; English.
 XX
 CC The invention relates to a secreted polypeptide comprising hepatitis C
 CC virus (HCV) E2 polypeptide lacking all or a portion of its membrane
 CC spanning domain so that the E2 polypeptide is capable of secretion into
 CC growth medium when expressed recombinantly in a host cell. The
 CC polypeptide may also lack a portion of its C-terminus. The HCV E2
 CC secreted polypeptide is useful for producing anti-HCV antibodies. A
 CC purified immunogenic polypeptide comprising HCV E2 is useful for treating
 CC HCV infection and for providing immune protection against HCV infection
 CC by administering it to a subject having or at risk of having HCV
 CC infection or in need of protection. The method further comprises
 CC administering an immunomodulatory agent such as interleukin-2 (IL-2),
 CC granulocyte macrophage-colony stimulating factor (GM-CSF) or
 CC gamma-interferon. The polypeptide is useful as a vaccine, and with other
 CC HCV proteins to form a multi-component HCV vaccine for prophylactic or
 CC therapeutic treatment of HCV infection. This sequence represents an N-
 CC and C-terminally truncated HCV delta-delta E2 genotype 1a protein.
 XX
 SQ Sequence 250 AA;
 Query Match 100.0%; Score 20; DB 23; Length 250;
 Best Local Similarity 100.0%; Pred. No. 2.1e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTDRSGAPTYSWG 20
 |||||
 DB 100 PSPVVVGTDRSGAPTYSWG 119

RESULT 14

ABP96704
 ID ABP96704 standard; protein; 251 AA.
 XX
 AC ABP96704;
 XX
 DT 04-JUN-2003 (first entry)
 XX
 DE E2 majority R9 amino acid sequence SEQ ID NO:2.
 XX
 KW Hepatitis C virus; HCV; E2 protein; E2 majority; vaccination; infection;
 KW hepatotropic; hepatitis C; vaccine; gene therapy; immunity;
 KW liver transplantation.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 4 /label= unknown
 FT /note= "encoded by TTT"
 XX
 PN WO2003022880-A2.
 XX
 PD 20-MAR-2003.
 XX
 PF 09-SEP-2002; 2002WO-IL00744.
 XX
 PR 13-SEP-2001; 2001IL-0145440.
 XX
 PA (XTLB-) XTL BIOPHARMACEUTICALS LTD.
 XX
 PI Neville L, Zauberman A;
 XX
 DR WPI; 2003-301052/29.
 DR N-PSDB; ACC44819.
 XX
 PT New purified and isolated E2 protein, useful for the manufacture of a
 PT vaccine composition for treating or preventing hepatitis C virus (HCV)
 PT infection, and for prophylactically inducing immunity against HCV -
 XX
 PS Claim 1; Page 32; 36pp; English.
 XX
 CC The present invention describes the modified Hepatitis C virus (HCV) E2
 CC proteins designated E2 majority, E2 majority R9 and E2 majority w/o (see
 CC ABP96703 to ABP96705), which are encoded by the nucleotide sequences
 CC given in ACC44818 to ACC44820. The modified E2 proteins have hepatotropic
 CC activities, and can be used in vaccines and in gene therapy. The modified
 CC E2 protein and compositions comprising them can be used for the
 CC manufacture of an HCV vaccine composition, for inducing immunity against
 CC HCV chronic carriers prior to, simultaneously to or after any other
 CC therapy, or in HCV-infected individuals prior to or after liver
 CC transplantation, or after presumed infection, and for prophylactically
 CC inducing immunity against HCV.
 XX
 SQ Sequence 251 AA;
 Query Match 100.0%; Score 20; DB 24; Length 251;
 Best Local Similarity 100.0%; Pred. No. 2.1e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTDRSGAPTYSWG 20
 |||||
 DB 104 PSPVVVGTDRSGAPTYSWG 120

RESULT 15
 AAB68043
 ID AAB68043 standard; protein; 254 AA.
 XX
 AC AAB68043;
 XX
 DT 29-JUN-2001 (first entry)
 XX
 XX

DE Amino acid sequence of water soluble variant of envelope E2 protein.
XX
KW Eo protein; HCV; envelope protein; E2 protein; HCV infection;
KW HCV attachment.
XX
OS Synthetic.
OS Hepatitis C virus.
XX
XX WO200122984-A1.
XX
XX PD 05-APR-2001.
XX
XX PF 26-SEP-2000; 2000WO-US26395.
XX
XX PR 29-SEP-1999; 99US-0407430.
XX
XX PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
XX PI Worman HJ, Mamiya N;
XX
XX DR WPI; 2001-273486/28.
XX
PT Treating or preventing hepatitis C virus infection in a subject,
PT involves administering hepatitis C virus envelope protein E2 binding
PT agents -
XX
PS Claim 5; Fig 8; 46pp; English.
XX
CC The present sequence represents a water soluble variant of a Hepatitis C
CC virus (HCV) envelope E2 protein. Agents that bind to the HCV E2 protein
CC (such as the human Eo protein), and so inhibit the attachment of HCV onto
CC cells (especially liver cells), are used to treat HCV infections in
CC mammals, in particular humans. The specification also describes a method
CC for identifying a compound which can be used for treating or preventing
CC HCV in a subject and which can inhibit the attachment of HCV onto cells
CC by inhibiting the binding of HCV envelope E2 protein to a cellular
CC protein associated with HCV attachment and entry into cells. The method
CC comprises incubating the compound, HCV envelope E2 protein or its variant
CC and a cellular protein capable of specifically binding to the HCV E2
CC protein under suitable reaction conditions; determining the interactions
CC between HCV envelope E2 protein and cellular protein in the presence and
CC absence of the compound; and comparing the interaction to identify a
CC compound which can inhibit the attachment of HCV onto cells.
XX
SQ Sequence 254 AA;

Query Match 100.0%; Score 20; DB 22; Length 254;
Best Local Similarity 100.0%; Pred. No. 2.1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
DB 106 PSPVVVGTTDRSGAPTYSWG 125

Search completed: November 21, 2003, 20:58:04
Job time : 32.15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.6 Seconds
(without alignments)
79.832 Million cell updates/sec

Title: US-09-973-025-81
Perfect score: 20
Sequence: 1 PSPVVVGTDRSGAPTYSWG 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0
Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-08-612-973-81
2	20	100.0	20	3	US-08-927-597-81
3	20	100.0	20	4	US-08-635-886C-37
4	20	100.0	179	3	US-08-444-818-77
5	20	100.0	305	3	US-08-478-073-2
6	20	100.0	333	1	US-08-453-552-12
7	20	100.0	333	2	US-08-710-637-12
8	20	100.0	333	5	PCT-US93-00907-12
9	20	100.0	337	1	US-08-188-281B-7
10	20	100.0	337	5	PCT-US94-07280-7
11	20	100.0	337	5	PCT-US95-01087-7
12	20	100.0	367	1	US-08-188-281B-9
13	20	100.0	367	1	US-08-453-552-6
14	20	100.0	367	2	US-08-710-637-6
15	20	100.0	367	5	PCT-US93-00907-6
16	20	100.0	367	5	PCT-US94-07280-9
17	20	100.0	367	5	PCT-US95-01087-9
18	20	100.0	377	1	US-08-188-281B-17
19	20	100.0	377	5	PCT-US94-07280-17
20	20	100.0	377	5	PCT-US95-01087-17
21	20	100.0	397	1	US-08-188-281B-11
22	20	100.0	397	5	PCT-US94-07280-11
23	20	100.0	397	5	PCT-US95-01087-11
24	20	100.0	399	1	US-08-453-552-10
25	20	100.0	399	2	US-08-710-637-10
26	20	100.0	399	5	PCT-US93-00907-10
27	20	100.0	402	1	US-08-460-806-13

28	20	100.0	402	1	US-08-460-806-15	Sequence 15, Appl
29	20	100.0	402	1	US-08-325-630-13	Sequence 13, Appl
30	20	100.0	402	1	US-08-325-630-15	Sequence 15, Appl
31	20	100.0	403	2	US-08-483-695-39	Sequence 39, Appl
32	20	100.0	403	2	US-07-985-285-39	Sequence 39, Appl
33	20	100.0	403	2	US-08-487-231-39	Sequence 39, Appl
34	20	100.0	403	3	US-09-201-912-39	Sequence 39, Appl
35	20	100.0	409	1	US-08-440-103-21	Sequence 21, Appl
36	20	100.0	409	1	US-08-440-103-24	Sequence 24, Appl
37	20	100.0	409	1	US-08-440-542-21	Sequence 21, Appl
38	20	100.0	409	1	US-08-440-542-24	Sequence 24, Appl
39	20	100.0	409	1	US-08-231-368-21	Sequence 21, Appl
40	20	100.0	409	1	US-08-231-368-24	Sequence 24, Appl
41	20	100.0	409	1	US-08-440-210-21	Sequence 21, Appl
42	20	100.0	409	1	US-08-440-210-24	Sequence 24, Appl
43	20	100.0	409	4	US-09-046-604-21	Sequence 21, Appl
44	20	100.0	409	4	US-09-046-604-24	Sequence 24, Appl
45	20	100.0	410	1	US-08-188-281B-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-08-612-973-81
; Sequence 81, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P. C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612.973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 81:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-612-973-81

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTDRSGAPTYSWG 20
|||||

Db 1 PSPVVVGTDRSGAPTYSWG 20

RESULT 2

US-08-927-597-81
; Sequence 81, Application US/08927597
; Patent No. 6245503
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 81:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-927-597-81

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTDRSGAPTYSWG 20

Db 1 PSPVVVGTDRSGAPTYSWG 20

RESULT 3

US-08-635-886C-37
; Sequence 37, Application US/08635886C
; Patent No. 655114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555

; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-37

Query Match 100.0%; Score 20; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTDRSGAPTYSWG 20

Db 1 PSPVVVGTDRSGAPTYSWG 20

RESULT 4

US-08-444-818-77
; Sequence 77, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 179 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-444-818-77

Query Match 100.0%; Score 20; DB 3; Length 179;
Best Local Similarity 100.0%; Pred. No. 7.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTDRSGAPTYSWG 20

Db 1 PSPVVVGTDRSGAPTYSWG 20

RESULT 5

US-08-478-073-2
; Sequence 2, Application US/08478073
; Patent No. 6020122
; GENERAL INFORMATION:
; APPLICANT: Okasinski, Gregory F.
; APPLICANT: Schaefer, Verlyn G.
; APPLICANT: Suhar, Thomas S.
; APPLICANT: Leeniewski, Richard R.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR NON-SECRETOR GENES
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: ONE HUNDRED ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,073
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Porembski, Priscilla E.
; REGISTRATION NUMBER: 33207
; REFERENCE/DOCKET NUMBER: 5763.US.01
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 305 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-478-073-2

Query Match 100.0%; Score 20; DB 3; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.1e-12; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

Qy 1 PSPVVVGTDRSGAPTYSWG 20
Db 152 PSPVVVGTDRSGAPTYSWG 171

RESULT 6
US-08-453-552-12
; Sequence 12, Application US/08453552
; Patent No. 5667992
; GENERAL INFORMATION:
; APPLICANT: CASEY, JAMES M.
; APPLICANT: BODE, SUZANNE L.
; APPLICANT: ZECK, BILLY J.
; APPLICANT: YAMAGUCHI, JULIE
; APPLICANT: FRAIL, DONALD E.
; APPLICANT: DESAI, SURESH M.
; APPLICANT: DEVARE, SUSHIL G.
; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,637
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/144,099
; FILING DATE:
; APPLICATION NUMBER: US 07/830,024
; FILING DATE: 01-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5131.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556

0
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,552
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5131.US.D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 333 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-453-552-12

Query Match 100.0%; Score 20; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.2e-12; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

Qy 1 PSPVVVGTDRSGAPTYSWG 20
Db 160 PSPVVVGTDRSGAPTYSWG 179

RESULT 7
US-08-710-637-12
; Sequence 12, Application US/08710637
; Patent No. 5854001
; GENERAL INFORMATION:
; APPLICANT: CASEY, JAMES M.
; APPLICANT: BODE, SUZANNE L.
; APPLICANT: ZECK, BILLY J.
; APPLICANT: YAMAGUCHI, JULIE
; APPLICANT: FRAIL, DONALD E.
; APPLICANT: DESAI, SURESH M.
; APPLICANT: DEVARE, SUSHIL G.
; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,637
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/144,099
; FILING DATE:
; APPLICATION NUMBER: US 07/830,024
; FILING DATE: 01-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5131.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556

; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 333 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-710-637-12

Query Match 100.0%; Score 20; DB 2; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.2e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
Db 160 PSPVVVGTTDRSGAPTYSWG 179

RESULT 8

PCT-US93-00907-12

; Sequence 12, Application PC/TUS9300907

; GENERAL INFORMATION:

; APPLICANT: CASEY, JAMES M.

; APPLICANT: BODE, SUZANNE L.

; APPLICANT: ZECK, BILLY J.

; APPLICANT: YAMAGUCHI, JULIE

; APPLICANT: FRAIL, DONALD E.

; APPLICANT: DESAI, SURESH M.

; APPLICANT: DEVARE, SUSHIL G.

; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D

; STREET: ONE ABBOTT PARK ROAD

; CITY: ABBOTT PARK

; STATE: IL

; COUNTRY: USA

; ZIP: 60064-3500

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US93/00907

; FILING DATE: 19930129

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: FOREMSKI, PRISCILLA E.

; REGISTRATION NUMBER: 33,207

; REFERENCE/DOCKET NUMBER: 5131.PC.01

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 708-937-6365

; TELEFAX: 708-937-9556

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 333 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: linear

; MOLECULE TYPE: protein

PCT-US93-00907-12

Query Match 100.0%; Score 20; DB 5; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.2e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
Db 160 PSPVVVGTTDRSGAPTYSWG 179

RESULT 9

US-08-188-281B-7

; Sequence 7, Application US/08188281B
; Patent No. 5610009
; GENERAL INFORMATION:
; APPLICANT: WATANABE, SHINICHI
; APPLICANT: YAMAGUCHI, JULIE
; APPLICANT: DESAI, SURESH M.
; APPLICANT: DEVARE, SUSHIL G.
; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/188,281B

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: FOREMSKI, PRISCILLA E.

REGISTRATION NUMBER: 33,207

REFERENCE/DOCKET NUMBER: 5521.US.01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 708-937-6365

TELEFAX: 708-938-2623

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 337 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-188-281B-7

Query Match 100.0%; Score 20; DB 1; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.2e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
Db 194 PSPVVVGTTDRSGAPTYSWG 213

RESULT 10

PCT-US94-07280-7

; Sequence 7, Application PC/TUS9407280

; GENERAL INFORMATION:

; APPLICANT: WATANABE, SHINICHI

; APPLICANT: YAMAGUCHI, JULIE

; APPLICANT: DESAI, SURESH M.

; APPLICANT: DEVARE, SUSHIL G.

; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D

; STREET: ONE ABBOTT PARK ROAD

; CITY: ABBOTT PARK

; STATE: IL

; COUNTRY: USA

; ZIP: 60064-3500

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US94/07280
;; FILING DATE:
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: FOREMSKI, PRISCILLA E.
;; REGISTRATION NUMBER: 33,207
;; REFERENCE/DOCKET NUMBER: 5521.US.01
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 708-937-6365
;; TELEFAX: 708-938-2623
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 337 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
PCT-US94-07280-7

Query Match 100.0%; Score 20; DB 5; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.2e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDRSGAPTYSWG 20
Db 194 PSPVVVGTTDRSGAPTYSWG 213

RESULT 11
PCT-US95-01087-7
;; Sequence 7, Application PC/TUS9501087
;; GENERAL INFORMATION:
;; APPLICANT: WATANABE, SHINICHI
;; APPLICANT: YAMAGUCHI, JULIE
;; APPLICANT: DESAI, SURESH M.
;; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
;; TITLE OF INVENTION: ENVELOPE GENES
;; NUMBER OF SEQUENCES: 22
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
;; STREET: ONE ABBOTT PARK ROAD
;; CITY: ABBOTT PARK
;; STATE: IL
;; COUNTRY: USA
;; ZIP: 60064-3500
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US95/01087
;; FILING DATE:
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: FOREMSKI, PRISCILLA E.
;; REGISTRATION NUMBER: 33,207
;; REFERENCE/DOCKET NUMBER: 5521.US.01
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 708-937-6365
;; TELEFAX: 708-938-2623
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 337 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
PCT-US95-01087-7

Query Match 100.0%; Score 20; DB 5; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.2e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDRSGAPTYSWG 20
Db 194 PSPVVVGTTDRSGAPTYSWG 213

RESULT 12
US-08-188-281B-9
;; Sequence 9, Application US/08188281B
;; Patent No. 5610009
;; GENERAL INFORMATION:
;; APPLICANT: WATANABE, SHINICHI
;; APPLICANT: YAMAGUCHI, JULIE
;; APPLICANT: DESAI, SURESH M.
;; APPLICANT: DEVARE, SUSHIL G.
;; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
;; TITLE OF INVENTION: ENVELOPE GENES
;; NUMBER OF SEQUENCES: 22
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
;; STREET: ONE ABBOTT PARK ROAD
;; CITY: ABBOTT PARK
;; STATE: IL
;; COUNTRY: USA
;; ZIP: 60064-3500
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/188,281B
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: FOREMSKI, PRISCILLA E.
;; REGISTRATION NUMBER: 33,207
;; REFERENCE/DOCKET NUMBER: 5521.US.01
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 708-937-6365
;; TELEFAX: 708-938-2623
;; INFORMATION FOR SEQ ID NO: 9:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 367 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-188-281B-9

Query Match 100.0%; Score 20; DB 1; Length 367;
Best Local Similarity 100.0%; Pred. No. 1.3e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDRSGAPTYSWG 20
Db 194 PSPVVVGTTDRSGAPTYSWG 213

RESULT 13
US-08-453-552-6
;; Sequence 6, Application US/08453552
;; Patent No. 5667992
;; GENERAL INFORMATION:
;; APPLICANT: CASEY, JAMES M.
;; APPLICANT: BODE, SUZANNE L.
;; APPLICANT: ZECK, BILLY J.
;; APPLICANT: YAMAGUCHI, JULIE
;; APPLICANT: FRAIL, DONALD E.
;; APPLICANT: DESAI, SURESH M.
;; APPLICANT: DEVARE, SUSHIL G.

;; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
;; TITLE OF INVENTION: PROTEINS
;; NUMBER OF SEQUENCES: 12
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
;; STREET: ONE ABBOTT PARK ROAD
;; CITY: ABBOTT PARK
;; STATE: IL
;; COUNTRY: USA
;; ZIP: 60064-3500
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/453,552
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: FOREMSKI, PRISCILLA E.
;; REGISTRATION NUMBER: 33,207
;; REFERENCE/DOCKET NUMBER: 5131.US.D1
;; TELEPHONE: 708-937-6365
;; TELEFAX: 708-937-9556
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 367 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-453-552-6

Query Match 100.0%; Score 20; DB 1; Length 367;
Best Local Similarity 100.0%; Pred. No. 1.3e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
Db 194 PSPVVVGTTDRSGAPTYSWG 213

RESULT 14
US-08-710-637-6
; Sequence 6, Application US/08710637
; Patent No. 5854001
; GENERAL INFORMATION:
; APPLICANT: CASEY, JAMES M.
; APPLICANT: BODE, SUZANNE L.
; APPLICANT: ZECK, BILLY J.
; APPLICANT: YAMAGUCHI, JULIE
; APPLICANT: FRAIL, DONALD E.
; APPLICANT: DESAI, SURESH M.
; APPLICANT: DEVARE, SUSHIL G.
; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,637
; FILING DATE:

;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/144,099
;; FILING DATE:
;; APPLICATION NUMBER: US 07/830,024
;; FILING DATE: 01-JAN-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: FOREMSKI, PRISCILLA E.
;; REGISTRATION NUMBER: 33,207
;; REFERENCE/DOCKET NUMBER: 5131.US.01
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 708-937-6365
;; TELEFAX: 708-937-9556
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 367 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-710-637-6

Query Match 100.0%; Score 20; DB 2; Length 367;
Best Local Similarity 100.0%; Pred. No. 1.3e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
Db 194 PSPVVVGTTDRSGAPTYSWG 213

RESULT 15
PCT-US93-00907-6
; Sequence 6, Application PC/TUS9300907
; GENERAL INFORMATION:
; APPLICANT: CASEY, JAMES M.
; APPLICANT: BODE, SUZANNE L.
; APPLICANT: ZECK, BILLY J.
; APPLICANT: YAMAGUCHI, JULIE
; APPLICANT: FRAIL, DONALD E.
; APPLICANT: DESAI, SURESH M.
; APPLICANT: DEVARE, SUSHIL G.
; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/00907
; FILING DATE: 19930129
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5131.PC.01
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 367 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein

PCT-US93-00907-6

Query Match 100.0%; Score 20; DB 5; Length 367;
Best Local Similarity 100.0%; Pred. No. 1.3e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTDRSGPTYSWG 20
|||
Db 194 PSPVVVGTDRSGPTYSWG 213
|||

Search completed: November 21, 2003, 21:15:16
Job time : 10.6 secs

0

0

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: November 21, 2003, 21:08:28 ; Search time 19.5 Seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-81
Perfect score: 20
Sequence: 1 PEPVVVGTTDRSGAPTSWG 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :		Published Applications AA:*	
1:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*		
2:	/cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*		
3:	/cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep:*		
4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*		
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*		
6:	/cgn2_6/ptodata/1/pubpaa/CTUS_PUBCOMB.pep:*		
7:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*		
8:	/cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*		
9:	/cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*		
10:	/cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*		
11:	/cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*		
12:	/cgn2_6/ptodata/1/pubpaa/US09D_PUBCOMB.pep:*		
13:	/cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*		
14:	/cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*		
15:	/cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*		
16:	/cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep:*		
17:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*		
18:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*		

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	10	US-09-973-025-81
2	20	100.0	20	11	US-09-899-303-81
3	20	100.0	20	11	US-09-995-808-81
4	20	100.0	20	11	US-09-995-860-81
5	20	100.0	20	12	US-09-995-791-81
6	20	100.0	28	10	US-09-921-397-6
7	20	100.0	31	10	US-09-921-397-15
8	20	100.0	37	10	US-09-921-397-4
9	20	100.0	43	10	US-09-921-397-22
10	20	100.0	96	10	US-09-921-397-82
11	20	100.0	176	10	US-09-921-397-81
12	20	100.0	250	10	US-09-952-572-8
13	20	100.0	254	10	US-09-407-430-3
14	20	100.0	350	10	US-09-929-955-4
15	20	100.0	350	14	US-10-104-966-4

16	20	100.0	363	10	US-09-407-430-2
17	20	100.0	363	12	US-10-128-587A-97
18	20	100.0	363	15	US-10-128-590-97
19	20	100.0	637	12	US-10-187-257-4
20	20	100.0	637	12	US-10-285-083-2
21	20	100.0	2894	10	US-09-941-611-23
22	20	100.0	2894	15	US-10-044-995-23
23	20	100.0	3011	9	US-09-742-659-4
24	20	100.0	3011	9	US-09-916-359-2
25	20	100.0	3011	10	US-09-238-076-20
26	20	100.0	3011	10	US-09-952-572-9
27	20	100.0	3011	10	US-09-929-955-1
28	20	100.0	3011	10	US-09-747-419-20
29	20	100.0	3011	11	US-09-891-894-3
30	20	100.0	3011	11	US-09-995-937-20
31	20	100.0	3011	11	US-09-917-563-20
32	20	100.0	3011	12	US-10-184-150-3
33	20	100.0	3011	14	US-10-104-966-1
34	20	100.0	3011	15	US-10-259-275-20
35	20	100.0	3011	16	US-10-232-643-6
36	20	100.0	3012	10	US-09-238-076-2
37	20	100.0	3012	11	US-09-995-937-2
38	20	100.0	3012	11	US-09-917-563-2
39	19	95.0	28	10	US-09-921-397-19
40	14	70.0	36	12	US-10-351-641-1084
41	11	55.0	31	16	US-10-318-200-13
42	11	55.0	290	12	US-10-128-587A-3
43	11	55.0	290	15	US-10-128-590-3
44	11	55.0	301	12	US-10-128-587A-5
45	11	55.0	301	15	US-10-128-590-5

ALIGNMENTS

RESULT 1

US-09-973-025-81
; Sequence 81, Application US/09973025
; Publication No. US20020182706A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 81:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 81:
US-09-973-025-81

Query Match 100.0%; Score 20; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
|||||

DB 1 PSPVVVGTTDRSGAPTYSWG 20
|||||

RESULT 2
US-09-899-303-81
; Sequence 81, Application US/09899303
; Publication No. US20030036110A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION NUMBER: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 81:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 81:
US-09-899-303-81

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
|||||

DB 1 PSPVVVGTTDRSGAPTYSWG 20

RESULT 3
US-09-995-808-81
; Sequence 81, Application US/09995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 81
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-81

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
|||||

DB 1 PSPVVVGTTDRSGAPTYSWG 20
|||||

RESULT 4
US-09-995-860-81
; Sequence 81, Application US/09995860
; Publication No. US20030118603A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and therapeutic use.
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 81
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-81

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
|||||

DB 1 PSPVVVGTTDRSGAPTYSWG 20
|||||

RESULT 5
US-09-995-793-81
; Sequence 81, Application US/09995791
; Publication No. US20030147918A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and therapeutic use.
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 81

; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-81

Query Match 100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
|||||

Db 1 PSPVVVGTTDRSGAPTYSWG 20
|||||

RESULT 6

US-09-921-397-6
; Sequence 6, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-6

Query Match 100.0%; Score 20; DB 10; Length 28;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
|||||

Db 1 PSPVVVGTTDRSGAPTYSWG 20
|||||

RESULT 7

US-09-921-397-15
; Sequence 15, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-15

Query Match 100.0%; Score 20; DB 10; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
|||||

Db 1 PSPVVVGTTDRSGAPTYSWG 20

RESULT 8

US-09-921-397-4
; Sequence 4, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-4

Query Match 100.0%; Score 20; DB 10; Length 37;
Best Local Similarity 100.0%; Pred. No. 1.9e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
|||||

Db 1 PSPVVVGTTDRSGAPTYSWG 20
|||||

RESULT 9

US-09-921-397-22
; Sequence 22, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 43
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-22

Query Match 100.0%; Score 20; DB 10; Length 43;
Best Local Similarity 100.0%; Pred. No. 2.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
|||||

Db 1 PSPVVVGTTDRSGAPTYSWG 36
|||||

RESULT 10

US-09-921-397-82
; Sequence 82, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and

```
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 82
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-82

Query Match      100.0%; Score 20; DB 10; Length 96;
Best Local Similarity 100.0%; Pred. No. 4.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
   |||||
Db 25 PSPVVVGTTDRSGAPTYSWG 44
   |||||

RESULT 11
US-09-921-397-81
; Sequence 81, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 81
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-81

Query Match      100.0%; Score 20; DB 10; Length 176;
Best Local Similarity 100.0%; Pred. No. 7.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
   |||||
Db 43 PSPVVVGTTDRSGAPTYSWG 62
   |||||

RESULT 12
US-09-952-572-8
; Sequence 8, Application US/09952572
; Patent No. US20020119495A1
; GENERAL INFORMATION:
; APPLICANT: HAWAII BIOTECHNOLOGY GROUP, Inc.
; APPLICANT: NAKANO, Eileen
; APPLICANT: CLEMENTS, David
; APPLICANT: HUMPHREYS, Tom
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITION OF HEPATITIS C
; FILE REFERENCE: HAWBIO1100
; CURRENT APPLICATION NUMBER: US/09/952,572
; CURRENT FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: US 60/230,927
; PRIOR FILING DATE: 2000-09-13
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Hepatitis C virus E2 protein sequence
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-09-929-955-4

Query Match      100.0%; Score 20; DB 10; Length 350;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-09-952-572-8

Query Match      100.0%; Score 20; DB 10; Length 250;
Best Local Similarity 100.0%; Pred. No. 1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
   |||||
Db 100 PSPVVVGTTDRSGAPTYSWG 119
   |||||

RESULT 13
US-09-407-430-3
; Sequence 3, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; TITLE OF INVENTION: HEPATITIS C VIRUS INFECTION
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-407-430-3

Query Match      100.0%; Score 20; DB 10; Length 254;
Best Local Similarity 100.0%; Pred. No. 1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
   |||||
Db 106 PSPVVVGTTDRSGAPTYSWG 125
   |||||

RESULT 14
US-09-929-955-4
; Sequence 4, Application US/09929955
; Patent No. US20020136740A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.23AUS2
; CURRENT APPLICATION NUMBER: US/09/929,955
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/225,767
; PRIOR FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-09-929-955-4

Query Match      100.0%; Score 20; DB 10; Length 350;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Wed Nov 26 17:25:35 2003

Qy 1 PSPVVVGTTDRSGPTYSWG 20
| | | | | | | | | | | | | | | | | |
Db 132 PSPVVVGTTDRSGPTYSWG 151

RESULT 15
US-10-104-966-4
; Sequence 4, Application US/10104966
; Publication No. US20020155124A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.23AUSC1
; CURRENT APPLICATION NUMBER: US/10/104,966
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-10-104-966-4

Query Match 100.0%; Score 20; DB 14; Length 350;
Best Local Similarity 100.0%; Pred No. 1.4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDRSGPTYSWG 20
| | | | | | | | | | | | | | | | | |
Db 132 PSPVVVGTTDRSGPTYSWG 151

Search completed: November 21, 2003, 22:19:39
Job time : 20.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 163.1 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-81

Perfect score: 20

Sequence: 1 PSPVVGTDRSCAPTYSWG 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_AA_Main.*
1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US097A_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US097B_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US099A_COMB.pep.*
25: /cgn2_6/ptodata/1/paa/US099B_COMB.pep.*
26: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
27: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
28: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
29: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
30: /cgn2_6/ptodata/1/paa/US104_COMB.pep.*
31: /cgn2_6/ptodata/1/paa/US106_COMB.pep.*
32: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	20	100.0	20 13	US-08-974-685-37
				Sequence 37, Appl

2	20	100.0	20 13	US-08-974-690-37	Sequence 37, Appl
3	20	100.0	20 13	US-08-974-690A-37	Sequence 37, Appl
4	20	100.0	20 13	US-08-974-690B-37	Sequence 37, Appl
5	20	100.0	20 13	US-08-974-690C-37	Sequence 37, Appl
6	20	100.0	20 23	US-09-899-303-81	Sequence 81, Appl
7	20	100.0	20 23	US-09-899-303A-81	Sequence 81, Appl
8	20	100.0	20 25	US-09-973-025-81	Sequence 81, Appl
9	20	100.0	20 25	US-09-995-791-81	Sequence 81, Appl
10	20	100.0	20 25	US-09-995-808-81	Sequence 81, Appl
11	20	100.0	20 25	US-09-995-860-81	Sequence 81, Appl
12	20	100.0	20 26	US-10-020-510-81	Sequence 81, Appl
13	20	100.0	20 29	US-10-321-798-81	Sequence 81, Appl
14	20	100.0	26 1	PCT-US02-17374-26	Sequence 26, Appl
15	20	100.0	28 24	US-09-921-397-6	Sequence 6, Appl
16	20	100.0	31 24	US-09-921-397-15	Sequence 15, Appl
17	20	100.0	37 24	US-09-921-397-4	Sequence 4, Appl
18	20	100.0	43 24	US-09-921-397-22	Sequence 22, Appl
19	20	100.0	96 24	US-09-921-397-82	Sequence 82, Appl
20	20	100.0	122 8	US-08-436-966-5	Sequence 5, Appl
21	20	100.0	176 24	US-09-921-397-81	Sequence 81, Appl
22	20	100.0	179 8	US-08-403-590B-77	Sequence 77, Appl
23	20	100.0	179 8	US-08-444-112-77	Sequence 77, Appl
24	20	100.0	250 1	PCT-US01-28767-8	Sequence 8, Appl
25	20	100.0	250 25	US-09-952-572-8	Sequence 8, Appl
26	20	100.0	254 1	PCT-US00-26395-3	Sequence 3, Appl
27	20	100.0	254 18	US-09-407-430-3	Sequence 3, Appl
28	20	100.0	333 8	US-08-417-478-12	Sequence 12, Appl
29	20	100.0	333 8	US-08-453-613-12	Sequence 12, Appl
30	20	100.0	333 30	US-10-445-724-3	Sequence 3, Appl
31	20	100.0	333 32	US-06-409-909-4	Sequence 4, Appl
32	20	100.0	350 21	US-09-705-547-4	Sequence 4, Appl
33	20	100.0	350 24	US-09-929-955-4	Sequence 4, Appl
34	20	100.0	350 27	US-10-104-966-4	Sequence 4, Appl
35	20	100.0	350 32	US-06-229-175-4	Sequence 4, Appl
36	20	100.0	363 1	PCT-US00-26395-2	Sequence 2, Appl
37	20	100.0	363 18	US-09-407-430-2	Sequence 2, Appl
38	20	100.0	363 27	US-10-128-587A-97	Sequence 97, Appl
39	20	100.0	363 27	US-10-128-590-97	Sequence 97, Appl
40	20	100.0	367 8	US-08-417-478-6	Sequence 6, Appl
41	20	100.0	367 8	US-08-453-613-6	Sequence 6, Appl
42	20	100.0	399 8	US-08-417-478-10	Sequence 10, Appl
43	20	100.0	399 8	US-08-453-613-10	Sequence 10, Appl
44	20	100.0	409 8	US-08-471-498-21	Sequence 21, Appl
45	20	100.0	409 8	US-08-471-498-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-08-974-685-37
; Sequence 37, Application US/08974685
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; DELEYS, ROBERT
; MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
; THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 181
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P. C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,685
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 37:
US-08-974-685-37

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.8e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDRSGAPTYSWG 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 PSPVVVGTTDRSGAPTYSWG 20

RESULT 2

US-08-974-690-37
Sequence 37, Application US/08974690
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 166
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P. C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690
FILING DATE: 25-APR-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: US 08/635,886
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

US-08-974-690-37

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.8e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDRSGAPTYSWG 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 PSPVVVGTTDRSGAPTYSWG 20

RESULT 3

US-08-974-690A-37
Sequence 37, Application US/08974690A
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
APPLICANT: MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P. C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690A
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 37:
US-08-974-690A-37

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.8e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDRSGAPTYSWG 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 PSPVVVGTTDRSGAPTYSWG 20

RESULT 4

US-08-974-690B-37
Sequence 37, Application US/08974690B
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
APPLICANT: MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303A
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 81:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 81:
US-09-899-303A-81
Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.8e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 PSPVVVGTDRSGAPTYSWG 20
Db 1 PSPVVVGTDRSGAPTYSWG 20
RESULT 8
US-09-973-025-81
; Sequence 81, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 81:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 81:
US-09-973-025-81
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.8e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 PSPVVVGTDRSGAPTYSWG 20
Db 1 PSPVVVGTDRSGAPTYSWG 20
RESULT 9
US-09-995-791-81
; Sequence 81, Application US/09995791
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 81
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-81
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.8e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 PSPVVVGTDRSGAPTYSWG 20
Db 1 PSPVVVGTDRSGAPTYSWG 20
RESULT 10
US-09-995-808-81
; Sequence 81, Application US/09995808
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 81
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-81
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.8e-12;

```
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PSPVVVGTTDRSGAPTYSWG 20
   |||||
Db 1 PSPVVVGTTDRSGAPTYSWG 20
   |||||

RESULT 11
US-09-995-860-81
; Sequence 81, Application US/09995860
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 81
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-81

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.8e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PSPVVVGTTDRSGAPTYSWG 20
   |||||
Db 1 PSPVVVGTTDRSGAPTYSWG 20
   |||||

RESULT 12
US-10-020-510-81
; Sequence 81, Application US/10020510
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-72
; CURRENT APPLICATION NUMBER: US/10/020,510
; CURRENT FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 81
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-020-510-81

Query Match 100.0%; Score 20; DB 26; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.8e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PSPVVVGTTDRSGAPTYSWG 20
   |||||
Db 1 PSPVVVGTTDRSGAPTYSWG 20
   |||||

RESULT 13
US-10-321-798-81
; Sequence 81, Application US/10321798
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-93
; CURRENT APPLICATION NUMBER: US/10/321,798
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: 60/418,358
; PRIOR FILING DATE: 2002-10-16
```

```
; PRIOR APPLICATION NUMBER: 10/020,510
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 81
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-321-798-81

Query Match 100.0%; Score 20; DB 29; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.8e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PSPVVVGTTDRSGAPTYSWG 20
   |||||
Db 1 PSPVVVGTTDRSGAPTYSWG 20
   |||||

RESULT 14
PCT-US02-17374-26
; Sequence 26, Application PC/TUS0217374
; GENERAL INFORMATION:
; APPLICANT: Acambis, Inc.
; APPLICANT: Kleanthous, Harold
; APPLICANT: Ors, Larissa
; APPLICANT: Miller, Charles
; TITLE OF INVENTION: Chimeric Flavivirus Vectors
; FILE REFERENCE: 06132/063W02
; CURRENT APPLICATION NUMBER: PCT/US02/17374
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/295,265
; PRIOR FILING DATE: 2001-06-01
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
PCT-US02-17374-26

Query Match 100.0%; Score 20; DB 1; Length 26;
Best Local Similarity 100.0%; Pred. No. 2.3e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PSPVVVGTTDRSGAPTYSWG 20
   |||||
Db 1 PSPVVVGTTDRSGAPTYSWG 20
   |||||

RESULT 15
US-09-921-397-6
; Sequence 6, Application US/09921397
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-6

Query Match 100.0%; Score 20; DB 24; Length 28;
Best Local Similarity 100.0%; Pred. No. 2.4e-12;
```

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTRSGAPTYSWG 20

Db 1 PSPVVVGTTRSGAPTYSWG 20

Search completed: November 21, 2003, 22:09:52
Job time : 163.1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.55 Seconds
(without alignments)
117.536 Million cell updates/sec

Title: US-09-973-025-81

Perfect score: 20

Sequence: 1 PSPVVVGTTDRSGAPTYSWG 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents AA New.*

- 1: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	6	US-10-651-165-37
2	20	100.0	30	6	US-10-296-734-476
3	20	100.0	333	1	PCT-US03-19834-3
4	20	100.0	333	6	US-10-655-562-4
5	20	100.0	637	1	PCT-US03-33610-4
6	20	100.0	2011	6	US-10-296-734-812
7	20	100.0	3011	1	PCT-US03-19834-2
8	20	100.0	3011	6	US-10-296-734-406
9	20	100.0	5985	6	US-10-296-734-810
10	17	85.0	2280	1	PCT-US03-20322-211
11	14	70.0	36	5	US-09-350-841A-1084
12	14	70.0	36	5	US-09-350-369D-1703
13	14	70.0	36	5	US-09-350-369E-1703
14	14	70.0	36	6	US-10-351-641-1084
15	13	65.0	30	6	US-10-296-734-474
16	13	65.0	2010	6	US-10-296-734-814
17	11	55.0	31	6	US-10-685-435-22
18	11	55.0	347	6	US-10-664-391-9
19	11	55.0	539	6	US-10-664-391-11
20	11	55.0	1026	1	PCT-US03-20409-3
21	11	55.0	3033	6	US-10-009-002-5
22	8	40.0	20	6	US-10-651-165-36
23	8	40.0	20	6	US-10-651-165-38
24	7	35.0	30	6	US-10-296-734-478
25	7	35.0	414	6	US-10-425-114A-56123
26	7	35.0	458	6	US-10-425-114A-41442

27	6	30.0	20	6	US-10-685-435-36	Sequence 36, Appl
28	6	30.0	46	5	US-09-897-516A-5212	Sequence 5212, Ap
29	78	5	78	5	US-09-897-516A-7100	Sequence 7100, Ap
30	6	30.0	113	5	US-09-897-516A-6024	Sequence 6024, Ap
31	6	30.0	115	5	US-09-897-516A-6466	Sequence 6466, Ap
32	6	30.0	163	6	US-10-385-415-96	Sequence 96, Appl
33	6	30.0	168	6	US-10-389-647-431	Sequence 431, App
34	6	30.0	186	6	US-10-425-114A-51742	Sequence 51742, A
35	6	30.0	189	6	US-10-425-114A-62348	Sequence 62348, A
36	6	30.0	195	6	US-10-425-114A-47868	Sequence 47868, A
37	6	30.0	238	6	US-10-425-114A-62167	Sequence 62167, A
38	6	30.0	249	6	US-10-425-114A-43163	Sequence 43163, A
39	6	30.0	266	6	US-10-679-063-6605	Sequence 6605, Ap
40	6	30.0	271	6	US-10-425-114A-37923	Sequence 37923, A
41	6	30.0	294	6	US-10-425-114A-62188	Sequence 62188, A
42	6	30.0	309	6	US-10-425-114A-70557	Sequence 70557, A
43	6	30.0	320	1	PCT-US03-26780-1464	Sequence 1464, Ap
44	6	30.0	320	1	PCT-US03-26780-1484	Sequence 1484, Ap
45	6	30.0	345	6	US-10-679-063-6039	Sequence 6039, Ap

ALIGNMENTS

RESULT 1

US-10-651-165-37
; Sequence 37, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651.165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974.690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-37

Query Match 100.0% Score 20; DB 6; Length 20;
Best Local Similarity 100.0%; Pred.No. 5.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDRSGAPTYSWG 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 PSPVVVGTTDRSGAPTYSWG 20

RESULT 2

US-10-296-734-476
; Sequence 476, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296.734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2

<p>; SEQ ID NO 476 ; LENGTH: 30 ; TYPE: PRT ; ORGANISM: Artificial ; FEATURE: ; OTHER INFORMATION: HepC 1a segment 35</p>	<p>Query Match 100.0%; Score 20; DB 6; Length 30; Best Local Similarity 100.0%; Pred. No. 7.5e-14; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>	<p>PCT-US03-19834-3 Sequence 3, Application PC/TUS0319834 GENERAL INFORMATION: APPLICANT: STAPLETON, JACK T. TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND METHODS FOR TREATING HYPERLIPIDEMIA FILE REFERENCE: IOWA:04SWO CURRENT APPLICATION NUMBER: PCT/US03/19834 PRIOR FILING DATE: 2003-06-24 PRIOR APPLICATION NUMBER: 10/445,724 PRIOR FILING DATE: 2003-05-27 PRIOR APPLICATION NUMBER: 60/392,158 PRIOR FILING DATE: 2002-06-28 NUMBER OF SEQ ID NOS: 3 SOFTWARE: PatentIn Ver. 2.1 SEQ ID NO 3 LENGTH: 333 TYPE: PRT ORGANISM: Hepatitis C virus PCT-US03-19834-3</p>	<p>Query Match 100.0%; Score 20; DB 1; Length 333; Best Local Similarity 100.0%; Pred. No. 5.7e-13; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>	<p>QY 1 PSPVVVGTTDRSGAPTYSWG 20 DB 129 PSPVVVGTTDRSGAPTYSWG 148</p>	<p>RESULT 3 PCT-US03-19834-3 Sequence 3, Application PC/TUS0319834 GENERAL INFORMATION: APPLICANT: STAPLETON, JACK T. TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND METHODS FOR TREATING HYPERLIPIDEMIA FILE REFERENCE: IOWA:04SWO CURRENT APPLICATION NUMBER: PCT/US03/19834 PRIOR FILING DATE: 2003-06-24 PRIOR APPLICATION NUMBER: 10/445,724 PRIOR FILING DATE: 2003-05-27 PRIOR APPLICATION NUMBER: 60/392,158 PRIOR FILING DATE: 2002-06-28 NUMBER OF SEQ ID NOS: 3 SOFTWARE: PatentIn Ver. 2.1 SEQ ID NO 3 LENGTH: 333 TYPE: PRT ORGANISM: Hepatitis C virus PCT-US03-19834-3</p>	<p>Query Match 100.0%; Score 20; DB 6; Length 333; Best Local Similarity 100.0%; Pred. No. 5.7e-13; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>	<p>QY 1 PSPVVVGTTDRSGAPTYSWG 20 DB 129 PSPVVVGTTDRSGAPTYSWG 148</p>	<p>RESULT 4 US-10-655-562-4 Sequence 4, Application US/10655562 GENERAL INFORMATION: APPLICANT: BISHOP-HURLEY, SHARON L. APPLICANT: SCHMIDT, FRANCIS J. APPLICANT: SMITH, ARNOLD L. TITLE OF INVENTION: PHASE-DISPLAY PEPTIDES AS NOVEL ANTIMICROBIAL AGENTS AGAINST HAEMOPHILUS INFLUENZAE FILE REFERENCE: UWMO:022US CURRENT APPLICATION NUMBER: US/10/655,562 CURRENT FILING DATE: 2003-09-04 PRIOR APPLICATION NUMBER: 60,409,909 PRIOR FILING DATE: 2002-09-11 NUMBER OF SEQ ID NOS: 8 SOFTWARE: PatentIn Ver. 2.1 SEQ ID NO 4 LENGTH: 333 TYPE: PRT ORGANISM: Artificial Sequence FEATURE: OTHER INFORMATION: Description of Artificial Sequence: Synthetic OTHER INFORMATION: Peptide US-10-655-562-4</p>	<p>Query Match 100.0%; Score 20; DB 6; Length 333; Best Local Similarity 100.0%; Pred. No. 5.7e-13; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>	<p>QY 1 PSPVVVGTTDRSGAPTYSWG 20 DB 129 PSPVVVGTTDRSGAPTYSWG 148</p>	<p>RESULT 5 PCT-US03-33610-4 Sequence 4, Application PC/TUS0333610 GENERAL INFORMATION: APPLICANT: HOUGHTON, Michael APPLICANT: COATES, Steve APPLICANT: SELBY, Mark APPLICANT: PALIARD, Xavier TITLE OF INVENTION: ACTIVATION OF HCV-SPECIFIC T-CELLS FILE REFERENCE: 2300-1612.60 CURRENT APPLICATION NUMBER: PCT/US03/33610 CURRENT FILING DATE: 2003-10-24 PRIOR APPLICATION NUMBER: 10/281,341 PRIOR FILING DATE: 2002-10-25 NUMBER OF SEQ ID NOS: 10 SOFTWARE: PatentIn Ver. 2.0 SEQ ID NO 4 LENGTH: 637 TYPE: PRT ORGANISM: Artificial Sequence FEATURE: OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region PCT-US03-33610-4</p>	<p>Query Match 100.0%; Score 20; DB 1; Length 637; Best Local Similarity 100.0%; Pred. No. 9.8e-13; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>	<p>QY 1 PSPVVVGTTDRSGAPTYSWG 20 DB 339 PSPVVVGTTDRSGAPTYSWG 358 0</p>	<p>RESULT 6 US-10-296-734-812 Sequence 812, Application US/10296734 GENERAL INFORMATION: APPLICANT: Thompson, Scott A APPLICANT: Ramshaw, Ian A TITLE OF INVENTION: Synthetic molecules and uses therefor FILE REFERENCE: Savine CURRENT APPLICATION NUMBER: US/10/296,734 CURRENT FILING DATE: 2003-08-04 PRIOR APPLICATION NUMBER: AU PQ7761/00 PRIOR FILING DATE: 2000-05-26 NUMBER OF SEQ ID NOS: 1507 SOFTWARE: Patentin version 3.2 SEQ ID NO 812 LENGTH: 2011 TYPE: PRT ORGANISM: Artificial FEATURE: OTHER INFORMATION: HepC cassette A US-10-296-734-812</p>	<p>Query Match 100.0%; Score 20; DB 6; Length 2011; Best Local Similarity 100.0%; Pred. No. 2.6e-12; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>	<p>QY 1 PSPVVVGTTDRSGAPTYSWG 20 DB 1324 PSPVVVGTTDRSGAPTYSWG 1343</p>
---	--	--	---	--	---	---	--	--	---	--	---	---	---	---	--	--

RESULT 7
PCT-US03-19834-2
; Sequence 2, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; TITLE OF INVENTION: METHODS FOR TREATING HYPERLIPIDEMIA
; FILE REFERENCE: IOWA:045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-2

Query Match 100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.6e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDRSGAPTYSWG 20
Db 511 PSPVVVGTTDRSGAPTYSWG 530

RESULT 8
US-10-296-734-406
; Sequence 406, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a consensus polypeptide
US-10-296-734-406

Query Match 100.0%; Score 20; DB 6; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.6e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDRSGAPTYSWG 20
Db 511 PSPVVVGTTDRSGAPTYSWG 530

RESULT 9
US-10-296-734-810
; Sequence 810, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734

; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 810
; LENGTH: 5985
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a savine
US-10-296-734-810

Query Match 100.0%; Score 20; DB 6; Length 5985;
Best Local Similarity 100.0%; Pred. No. 6.4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDRSGAPTYSWG 20
Db 1323 PSPVVVGTTDRSGAPTYSWG 1342

RESULT 10
PCT-US03-20322-211
; Sequence 211, Application PC/TUS0320322
; GENERAL INFORMATION:
; APPLICANT: Dana-Farber Cancer Institute, Inc., et al.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
; FILE REFERENCE: DFN-043PC
; CURRENT APPLICATION NUMBER: PCT/US03/20322
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 211
; LENGTH: 2280
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
PCT-US03-20322-211

Query Match 85.0%; Score 17; DB 1; Length 2280;
Best Local Similarity 100.0%; Pred. No. 2.9e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDRSGAPT Y 17
Db 511 PSPVVVGTTDRSGAPT Y 527

RESULT 11
US-09-350-841A-1084
; Sequence 1084, Application US/09350841A
; GENERAL INFORMATION:
; APPLICANT: Jeffes, Peter;
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING HIV TRANSMISSION
; FILE REFERENCE: 7872-066-999
; CURRENT APPLICATION NUMBER: US/09/350,841A
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 1946
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1084
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-09-350-841A-1084

Query Match 70.0%; Score 14; DB 5; Length 36;
Best Local Similarity 100.0%; Pred. No. 9.3e-08;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GTDRSGAPTYSWG 20
| | | | | | | | | |
Db 2 GTDRSGAPTYSWG 15

RESULT 12

US-09-350-369D-1703
; Sequence 1703, Application US/09350369D
; GENERAL INFORMATION:
; APPLICANT: Samuel E. Hopkins
; APPLICANT: Elizabeth M. Dimassimo
; APPLICANT: Pamela P. Runak
; APPLICANT: Thomas Michael Venetta
; TITLE OF INVENTION: Methods and Compositions for Detection of HIV GP41 Peptides and H
; FILE REFERENCE: 7872-068-999
; CURRENT APPLICATION NUMBER: US/09/350,369D
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 2088
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1703
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus-1 (HIV-1)
US-09-350-369D-1703

Query Match 70.0%; Score 14; DB 5; Length 36;
Best Local Similarity 100.0%; Pred. No. 9.3e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GTDRSGAPTYSWG 20
| | | | | | | | | |
Db 2 GTDRSGAPTYSWG 15

RESULT 13

US-09-350-369E-1703
; Sequence 1703, Application US/09350369E
; GENERAL INFORMATION:
; APPLICANT: Samuel E. Hopkins
; APPLICANT: Elizabeth M. Dimassimo
; APPLICANT: Pamela P. Runak
; APPLICANT: Thomas Michael Venetta
; TITLE OF INVENTION: Methods and Compositions for Detection of HIV GP41 Peptides and H
; FILE REFERENCE: 7872-068-999
; CURRENT APPLICATION NUMBER: US/09/350,369E
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 2088
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1703
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus-1 (HIV-1)
US-09-350-369E-1703

Query Match 70.0%; Score 14; DB 5; Length 36;
Best Local Similarity 100.0%; Pred. No. 9.3e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GTDRSGAPTYSWG 20
| | | | | | | | | |
Db 2 GTDRSGAPTYSWG 15

RESULT 14

US-10-351-641-1084
; Sequence 1084, Application US/10351641
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.

; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
; APPLICANT: Lambert, D.
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; TITLE OF INVENTION: PROPERTIES
; FILE REFERENCE: 7872-100
; CURRENT APPLICATION NUMBER: US/10/351,641
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: 09/350,641
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/315,304
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1757
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1084
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-10-351-641-1084

Query Match 70.0%; Score 14; DB 6; Length 36;
Best Local Similarity 100.0%; Pred. No. 9.3e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GTDRSGAPTYSWG 20
| | | | | | | | | |
Db 2 GTDRSGAPTYSWG 15

RESULT 15

US-10-296-734-474
; Sequence 474, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 474
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 34
US-10-296-734-474

Query Match 65.0%; Score 13; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 8.1e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSG 13
| | | | | | | | | |
Db 18 PSPVVVGTTDRSG 30

Search completed: November 21, 2003, 22:12:56
Job time : 8.55 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:46:27 ; Search time 9.25 Seconds
(without alignments)
207.932 Million cell updates/sec

Title: US-09-973-025-81

Perfect score: 20

Sequence: 1 PSPVVVGTTDRSGAPTYSWG 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 9616862 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR 76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	3011	1 GNVVC3	genome polyprotein
2	20	100.0	3011	1 GNVVCH	genome polyprotein
3	17	85.0	640	2 JQ1584	genome polyprotein
4	17	85.0	782	2 S19875	genome polyprotein
5	17	85.0	3010	1 A45573	genome polyprotein
6	17	85.0	3011	1 S40770	genome polyprotein
7	13	65.0	415	2 PC4407	envelope protein -
8	11	55.0	234	2 S32742	genome polyprotein
9	11	55.0	235	2 S32747	genome polyprotein
10	11	55.0	237	2 S32744	genome polyprotein
11	11	55.0	523	2 JQ1926	polyprotein - hepa
12	11	55.0	782	2 S19876	genome polyprotein
13	11	55.0	782	2 S18031	genome polyprotein
14	11	55.0	782	2 S18032	genome polyprotein
15	11	55.0	787	2 PN0677	hypothetical prote
16	11	55.0	3010	1 GNVVTC	genome polyprotein
17	11	55.0	3010	1 GNVVCH	genome polyprotein
18	11	55.0	3010	1 S18030	genome polyprotein
19	11	55.0	3010	1 GNVVTV	genome polyprotein
20	11	55.0	3014	1 JCS620	genome polyprotein
21	11	55.0	3033	1 JQ1303	genome polyprotein
22	10	50.0	876	2 PC2219	polypeptide - hepa
23	10	50.0	3033	1 GNVVJ8	genome polyprotein
24	9	45.0	520	2 JQ1925	polyprotein - hepa
25	9	45.0	716	2 JQ1366	polyprotein - hepa
26	6	30.0	82	1 VL8K1	apovitellenin I -
27	6	30.0	95	1 KRDKF4	keratin B-4, feath
28	6	30.0	95	1 KRPFY4	keratin B-4, feath
29	6	30.0	98	1 KRCHF1	keratin I, feather

RESULT 1 0 6 30.0 98 1 KRCHF2
GNWVC3
genome polyprotein - hepatitis C virus (strain HCV-1)
N:Contains: capsid protein C; envelope protein M; hepatitisvirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 19-Jan-2001
C:Accession: A39166; PQ0403; PQ0404
R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Co-
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A:Title: Genetic organization and diversity of the hepatitis C virus.
A:Reference number: A39166; MUID:91172826; PMID:1848704
A:Accession: A39166
A:Molecule type: mRNA
A:Residues: 1-3011 <CHO>
A:Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874
J. Gen. Virol. 73, 1131-1141, 1992
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to
A:Reference number: PQ0393; MUID:92268871; PMID:1316939
A:Accession: PQ0403
A:Molecule type: Genomic RNA
A:Residues: 1577-1633 <CHA>
A:Cross-references: DBJ:D10128
A:Experimental source: isolates E-b16
A:Accession: PQ0404
A:Status: preliminary
A:Molecule type: Genomic RNA
A:Residues: 1577-1633 <CH2>
A:Experimental source: isolates E-b17
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F:1-115/Product: capsid protein C #status predicted <CP>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <ME>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitisvirin #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
F:1863-2011/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2034-3013/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 100.0% Score 20; DB 1; Length 3011;

Best Local Similarity 100.0%; Pred. No. 2.1e-12;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20

|||||

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 85.0%; Score 17; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 2.5e-09; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 0;

Qy 1 PSPVVVGTDRSGAPTY 17
|||||
Db 511 PSPVVVGTDRSGAPTY 527

RESULT 6

S40770
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (fragment)
N:Contains: envelope protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: S40770; PC1285
R:Okamoto, H.
submitted to the EMBL Data Library, March 1992
A:Reference number: S40770
A:Accession: S40770
A:Molecule type: genomic RNA
A:Residues: 1-3011 <OKA>
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; PMID:91013116; PMID:2170712
A:Accession: PC1285
A:Molecule type: genomic RNA
A:Residues: 1-513 <OK2>
A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
A:Experimental source: isolate HC-J1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 85.0%; Score 17; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.5e-09; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 0;

Qy 1 PSPVVVGTDRSGAPTY 17
|||||
Db 511 PSPVVVGTDRSGAPTY 527

RESULT 7

PC4407
envelope protein - hepatitis C virus (fragment)
C:Species: hepatitis C virus
C:Date: 10-Nov-1997 #sequence_revision 23-Jan-1998 #text_change 17-Nov-2000
C:Accession: PC4407
R:Li, G.; Yao, J.; Peng, W.
Chinese J. Virol. 13, 24-32, 1997
A:Title: Sequence of genomic region of hepatitis C virus envelope proteins from a Guang
A:Reference number: PC4407
A:Accession: PC4407
A:Molecule type: genomic RNA
A:Residues: 1-415 <LIA>

A:Note: the authors translated the codon ATA for residues 93 and 249 as Met
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein

Query Match 65.0%; Score 13; DB 2; Length 415;
Best Local Similarity 100.0%; Pred. No. 5.3e-06; Indels 0; Gaps 0;
Matches 13; Conservative 0; Mismatches 0;

Qy 1 PSPVVVGTDRSG 13
|||||
Db 346 PSPVVVGTDRSG 358

RESULT 8

S32742
genome polyprotein - hepatitis C virus (isolate CR-1) (fragment)
N:Contains: envelope protein E2
C:Species: hepatitis C virus
A:Variety: isolate CR-1
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S32742
R:Roggendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
submitted to the EMBL Data Library, February 1993
A:Description: Variability of the envelope regions of HCV in European isolates and its
A:Reference number: S32741
A:Accession: S32742
A:Molecule type: genomic RNA
A:Residues: 1-234 <ROG>
A:Cross-references: EMBL:X72979; NID:g296102; PIDN:CAA51484.1; PID:g296103
A:Experimental source: isolate Cr-1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; polyprotein
F:1-234/Product: envelope protein E2 #status predicted <MAT>

Query Match 55.0%; Score 11; DB 2; Length 234;
Best Local Similarity 100.0%; Pred. No. 0.00036; Indels 0; Gaps 0;
Matches 11; Conservative 0; Mismatches 0;

Qy 1 PSPVVVGTDR 11
|||||
Db 142 PSPVVVGTDR 152

RESULT 9

S32747
genome polyprotein - hepatitis C virus (isolate HU-1) (fragment)
N:Contains: envelope protein E2
C:Species: hepatitis C virus
A:Variety: isolate HU-1
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S32747
R:Roggendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
submitted to the EMBL Data Library, February 1993
A:Description: Variability of the envelope regions of HCV in European isolates and its
A:Reference number: S32741
A:Accession: S32747
A:Molecule type: genomic RNA
A:Residues: 1-235 <ROG>
A:Cross-references: EMBL:X72977; NID:g296112; PIDN:CAA51482.1; PID:g296113
A:Experimental source: isolate HU-1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; polyprotein
F:1-235/Product: envelope protein E2 #status predicted <MAT>

Query Match 55.0%; Score 11; DB 2; Length 235;
Best Local Similarity 100.0%; Pred. No. 0.00036; Indels 0; Gaps 0;
Matches 11; Conservative 0; Mismatches 0;

Qy 1 PSPVVVGTDR 11
|||||
Db 140 PSPVVVGTDR 153

```
RESULT 10
S32744
genome polyprotein - hepatitis C virus (isolate EG-1) (fragment)
N:Contains: envelope protein E2
C:Species: hepatitis C virus
A:Variety: isolate EG-1
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S32744
R:Roggendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
submitted to the EMBL Data Library, February 1993
A:Description: Variability of the envelope regions of HCV in European isolates and its s
A:Reference number: S32741
A:Accession: S32744
A:Molecule type: genomic RNA
A:Residues: 1-237 <ROG>
A:Cross-references: EMBL:X72981
A:Experimental source: isolate EG-1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; polyprotein
F:1-237/Product: envelope protein E2 #status predicted <MAT>

Query Match 55.0%; Score 11; DB 2; Length 237;
Best Local Similarity 100.0%; Pred. No. 0.00036;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDR 11
||| ||||| |||||
Db 145 PSPVVVGTTDR 155

RESULT 11
JQ1926
polyprotein - hepatitis C virus (isolate HCV-476)
N:Contains: C protein; E1 protein; E2/NS1 protein
C:Species: hepatitis C virus
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 17-Nov-2000
C:Accession: JQ1926
R:Abe, K.; Inchauspe, G.; Fujisawa, K.
J. Gen. Virol. 73, 2725-2729, 1992
A:Title: Genomic characterization and mutation rate of hepatitis C virus isolated from a
A:Reference number: JQ1925; MUID:93019030; PMID:1383400
A:Accession: JQ1926
A:Molecule type: mRNA
A:Residues: 1-523 <ABE>
A:Cross-references: DBJ:DJ10687
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein
F:1-191/Product: C protein #status predicted <CPR>
F:192-383/Product: E1 protein #status predicted <E1R>
F:384-523/Product: E2/NS1 protein #status predicted <E2P>

Query Match 55.0%; Score 11; DB 2; Length 523;
Best Local Similarity 100.0%; Pred. No. 0.00074;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDR 11
||| ||||| |||||
Db 512 PSPVVVGTTDR 522

RESULT 12
S19876
genome polyprotein - hepatitis C virus (isolate JK5) (fragment)
N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C:Species: hepatitis C virus
A:Variety: isolate JK5
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C:Accession: S19876
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
A:Reference number: S18029
A:Accession: S19876
```

```
A:Molecule type: genomic RNA
A:Residues: 1-782 <HON>
A:Cross-references: EMBL:X61595; NID:g59486; PIDN:CAA43792.1; PID:g59487
A:Experimental source: isolate JK5
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F:1-191/Product: core protein #status predicted <MAT1>
F:192-383/Product: envelope protein 1 #status predicted <MAT2>
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 55.0%; Score 11; DB 2; Length 782;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDR 11
||| ||||| |||||
Db 511 PSPVVVGTTDR 521

RESULT 13
S18031
genome polyprotein - hepatitis C virus (isolate JK2) (fragment)
N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C:Species: hepatitis C virus
A:Variety: isolate JK2
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C:Accession: S18031
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
A:Reference number: S18029
A:Accession: S18031
A:Molecule type: genomic RNA
A:Residues: 1-782 <HON>
A:Cross-references: EMBL:X61593
A:Experimental source: isolate JK2
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F:1-191/Product: core protein #status predicted <MAT1>
F:192-383/Product: envelope protein 1 #status predicted <MAT2>
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 55.0%; Score 11; DB 2; Length 782;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDR 11
||| ||||| |||||
Db 511 PSPVVVGTTDR 521

RESULT 14
S18032
genome polyprotein - hepatitis C virus (isolate JK4) (fragment)
N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C:Species: hepatitis C virus
A:Variety: isolate JK4
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C:Accession: S18032
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
A:Reference number: S18029
A:Accession: S18032
A:Molecule type: genomic RNA
A:Residues: 1-782 <HON>
A:Cross-references: EMBL:X61594
A:Experimental source: isolate JK4
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F:1-191/Product: core protein #status predicted <MAT1>
```

F:192-383/Product: envelope protein 1 #status predicted <MAT2>
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 55.0%; Score 11; DB 2; Length 782;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDR 11
|||
Db 511 PSPVVVGTTDR 521

RESULT 15
PN0677
hypothetical protein 787 - hepatitis C virus (fragment)
C:Species: hepatitis C virus
C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000
C:Accession: PN0677
R:Cho, S.H.; Yoon, J.I.; Chang, J.E.; Ahn, B.M.; Lee, C.H.; Lee, Y.I.
Biochem. Biophys. Res. Commun. 196, 780-788, 1993
A:Title: Genomic typing of hepatitis C viruses from Korean patients: Implications of gen
A:Reference number: PN0677; MUID:94059104; PMID:8240354
A:Accession: PN0677
A:Molecule type: mRNA
A:Residues: 1-787 <CHO>
A:Cross-references: GB:L20498; NID:g1381031; PIDN:AAB02608.1; PID:g1381032
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: glycoprotein; nonstructural protein
F:196,209,234,250,305,325,421,427,452,536,544,560,580,627,649/Binding site: carbohydrate

Query Match 55.0%; Score 11; DB 2; Length 782;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDR 11
|||
Db 515 PSPVVVGTTDR 525

Search completed: November 21, 2003, 21:11:32
Job time : 9.25 secs

THIS PAGE BLANK (USPTO)

DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RGRP.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR007095; RNA pol DS PS.
 DR InterPro; IPR007094; RNA pol psvir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV env; 1.
 DR Pfam; PF01539; HCV NS1; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; Helicase C; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 393 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 394 729 MAJOR ENVELOPE PROTEIN E1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 136 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 1.6e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PSPVVVGTDRSGAPTYSWG 20

Db 511 PSPVVVGTDRSGAPTYSWG 530
 RESULT 2
 POLG_HCVH STANDARD; PRT; 3011 AA.
 ID POLG_HCVH STANDARD; PRT; 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV)
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=1658800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.B., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding";
 RL Structure 6:89-100(1998).
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position. Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.

CC DR EMBL; M67463; AAA45534.1; -
DR PIR; A36814; GNVVCH.
DR PDB; 1HEI; 25-NOV-98.
DR PDB; 1AIV; 16-PEB-99.
DR PDB; 1AIR; 17-JUN-98.
DR MEROPS; S29.001; -.
DR TRANSFAC; T04155; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA pol DS ps.
DR InterPro; IPR007094; RNA pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02307; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
FT CHAIN 192 383 CAPSID PROTEIN C.
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
FT CHAIN 810 1026 PROTEIN P7.
FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5A.
FT CHAIN 3011 369 NONSTRUCTURAL PROTEIN NS5B.
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DCH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT STRAND 1224 1226
FT TURN 1232 1233
FT TURN 1236 1238
FT HELIX 1239 1246
FT TURN 1247 1248
FT STRAND 1251 1255
FT HELIX 1258 1271
FT TURN 1272 1280
FT TURN 1281 1282
FT STRAND 1283 1285
FT STRAND 1291 1295
FT HELIX 1296 1301
FT STRAND 1302 1303
FT TURN 1312 1316
FT TURN 1317 1319
FT HELIX 1323 1335
FT TURN 1336 1340
FT STRAND 1343 1347
FT TURN 1352 1353
FT TURN 1360 1361
FT STRAND 1362 1366
FT STRAND 1368 1368
FT STRAND 1373 1375
FT TURN 1376 1377
FT STRAND 1378 1380
FT HELIX 1382 1385
FT STRAND 1389 1393
FT HELIX 1397 1409
FT TURN 1410 1411
FT STRAND 1414 1417
FT TURN 1419 1420
FT STRAND 1432 1436
FT TURN 1438 1439
FT STRAND 1450 1453
FT STRAND 1456 1463
FT STRAND 1471 1478
FT STRAND 1480 1480
FT HELIX 1481 1488
FT TURN 1489 1490
FT STRAND 1497 1501
FT STRAND 1507 1507
FT STRAND 1511 1511
FT HELIX 1514 1527
FT HELIX 1532 1544
FT STRAND 1550 1550
FT HELIX 1555 1564
FT HELIX 1570 1578
FT TURN 1579 1580
FT HELIX 1584 1597
FT TURN 1598 1598
FT TURN 1606 1611
FT TURN 1614 1618
FT STRAND 1622 1623
FT STRAND 1627 1627
FT STRAND 1635 1636
FT HELIX 1640 1652
SQ SEQUENCE 3011 AA; 772CBB29CCD94753 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.6e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTDRSGAPTSWG 20
|||||
Db 511 PSPVVVGTDRSGAPTSWG 530
|||||

RESULT 3
POLG_HCVUT
ID POLG_HCVUT STANDARD; PRT; 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J7) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92295714; PubMed=1318627;
 RA Tanaka T., Kato N., Nakagawa M., Otsuyama Y., Cho M.J.,
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 RA "Molecular cloning of hepatitis C virus genome from a single Japanese
 RT carrier: sequence variation within the same individual and among
 RT infected individuals.";
 RL Virus Res. 23:39-53(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D11168; BAA01943.1; --
 CC PIR; A45573; A45573.
 CC PDB; 1AIQ; 25-MAR-98.
 CC PDB; 1JXP; 14-JAN-98.
 CC MEROPS; S29.001; --
 CC MEROPS; U39.001; --
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4b.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02307; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.

DR Pfam; PF00998; Viral RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDc_1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;
 Query Match 85.0%; Score 17; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 1.7e-09;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PSPVVGTTDRSGAPTY 17
 DB 51b PSPVVGTTDRSGAPTY 527
 RESULT 4
 POLG_HCVJ7
 ID POLG_HCVJ7 STANDARD; PRT; 737 AA.
 AC P27961;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
 DE protein (Envelope protein M); Major envelope protein E; Nonstructural
 DE proteins NS1 and NS2] (fragment).
 OS Hepatitis C virus (isolate HC-J7) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11114;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,

RA Tanaka T., Fukuda S., Teuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D10077; BAA00971.1; -;
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01541; HCV env; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR PolyProtein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 737
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 233
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 430
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT NON_TER 737 737
 SQ SEQUENCE 737 AA; 81691 MW; 67DPAE11854122P2 CRC64;

Query Match 55.0%; Score 11; DB 1; Length 737;
 Best Local Similarity 100.0%; Pred. No. 0.00061;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPSVVGTTDR 11
 |||||
 DB 513 PPSVVGTTDR 523

RESULT 5
 POLG-HCVBK STANDARD; PRT; 3010 AA.
 AC P26663;

DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate BK) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=111105;
 OX [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91140698; PubMed=1847440;
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 RA Onishi E., Andoh T., Yoshida I., Okayama H.;
 RT "Structure and organization of the hepatitis C virus genome isolated
 RT from human carriers.";
 RL J. Virol. 65:1105-1113 (1991).
 RN [2]
 RP SEQUENCE OF 1487-1500.
 RX MEDLINE=96235224; PubMed=8647104;
 RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetky L.;
 RT "Non-structural protein 3 of hepatitis C virus inhibits
 RT phosphorylation mediated by cAMP-dependent protein kinase.";
 RL Eur. J. Biochem. 237:611-618(1996).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RX MEDLINE=97015088; PubMed=8861916;
 RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
 RA Moomaw E.W., Adachi T., Hostonska Z.;
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 RT trypsin-like fold and a structural zinc binding site.";
 RL Cell 87:331-342(1996).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RX MEDLINE=98227846; PubMed=9568891;
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
 RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
 RL Protein Sci. 7:837-847(1998).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC [RNA] (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M58335; AAA72945.1; -;
 DR PIR; A38465; GNVVTC
 DR PDB; 1AIQ; 25-MAR-98.
 DR PDB; 1UXP; 14-JAN-98.
 DR PDB; 1NS3; 08-APR-98.
 DR PDB; 1C2P; 15-NOV-00.

DR PDB: 1CSJ; 08-NOV-99.
 DR PDB: 1GX5; 09-APR-02.
 DR PDB: 1GX6; 10-APR-02.
 DR PDB: 1QUV; 26-JUN-00.
 DR PDB: 8OHM; 20-APR-99.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002531; HCV NS2.
 DR InterPro; IPR002518; HCV NS3.
 DR InterPro; IPR004109; HCV NS4.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR007095; RNA pol DS PS.
 DR InterPro; IPR007094; RNA pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).
 FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 3010 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 1396 1396 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT STRAND 1031 1035
 FT HELIX 1039 1047
 FT STRAND 1050 1050
 FT STRAND 1059 1063
 FT STRAND 1068 1074
 FT TURN 1075 1076
 FT STRAND 1077 1081
 FT HELIX 1084 1085
 FT TURN 1086 1087
 FT STRAND 1090 1092
 FT TURN 1093 1094
 FT STRAND 1095 1097
 FT STRAND 1101 1103
 FT TURN 1104 1107
 FT STRAND 1108 1112
 FT STRAND 1120 1120
 FT STRAND 1122 1122
 FT STRAND 1129 1133
 FT TURN 1135 1136
 FT STRAND 1139 1144
 FT STRAND 1149 1157
 FT HELIX 1158 1161
 FT TURN 1162 1163
 FT STRAND 1165 1166
 FT STRAND 1168 1171
 FT TURN 1172 1174
 FT STRAND 1175 1186
 FT TURN 1187 1188
 FT STRAND 1189 1197
 FT HELIX 1198 1202
 FT TURN 1203 1204
 FT STRAND 1680 1688
 SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCFDFD9C CRC64;
 Query Match 55.0%; Score 11; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.0021;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 PSPVVVGTTDR 11
 Db 511 PSPVVVGTTDR 521
 RESULT 6
 POLG_HCVJA *STANDARD; PRT; 3010 AA.
 ID POLG_HCVJA *STANDARD;
 AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Japanese) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=91089550; PubMed=2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 RA Sugimura T., Shimotohno K.;
 RT "Molecular cloning of the human hepatitis C virus genome from
 RT Japanese patients with non-A, non-B hepatitis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RN [2]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=91192160; PubMed=1849488;
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K.,

Ohkoshi S., Shimotohno K.;
 "Molecular structure of the Japanese hepatitis C viral genome.";
 FEBS Lett. 280:325-328(1991).
 -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 (RNA) (N).
 -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA.
 -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; D90208; BAA14233.1; -;
 PIR; A39253; GNMVCJ.
 HSRP; P26663; 1JXP.
 MEROPS; S29.001; -;
 MEROPS; U39.001; -;
 InterPro; IPR001410; DEAD.
 InterPro; IPR002522; HCV capsid.
 InterPro; IPR002521; HCV core.
 InterPro; IPR002519; HCV env.
 InterPro; IPR002531; HCV NS1.
 InterPro; IPR002518; HCV NS2.
 InterPro; IPR004109; HCV NS3.
 InterPro; IPR000745; HCV NS4a.
 InterPro; IPR001490; HCV NS4b.
 InterPro; IPR002868; HCV NS5a.
 InterPro; IPR002166; HCV NS5a.
 InterPro; IPR001850; Helicase_C.
 InterPro; IPR007095; RNA pol_PS.
 InterPro; IPR007094; RNA pol_PSvir.
 Pfam; PF01543; HCV capsid; 1.
 Pfam; PF01542; HCV core; 1.
 Pfam; PF01539; HCV env; 1.
 Pfam; PF01560; HCV NS1; 1.
 Pfam; PF01538; HCV NS2; 1.
 Pfam; PF02907; HCV NS3; 1.
 Pfam; PF01006; HCV NS4a; 1.
 Pfam; PF01001; HCV NS4b; 1.
 Pfam; PF01506; HCV NS5a; 1.
 Pfam; PF00271; Helicase_C; 1.
 Pfam; PF00998; Viral RdRp; 1.
 ProDom; PD186062; HCV NS1; 1.
 SMART; SM00487; DEXoc; 1.
 Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 Transmembrane; Nonstructural
 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 INIT_MET 1 1
 CHAIN 1 115
 CHAIN 116 191
 CHAIN 192 383
 CHAIN 384 729
 CHAIN 730 1006
 CHAIN 1007 1615
 CHAIN 1616 1862
 CHAIN 1863 2013
 CHAIN 2014 3010
 CHAIN 3011 369
 TRANSMEM 347 369
 ACT_SITE 1083 1083
 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT SITE 1107 1107
 FT NP_BIND 1165 1165
 FT SITE 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;
 Query Match 55.0%; Score 11; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.0021;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PSPVVVGTTDR 11
 DB 511 PSPVVVGTTDR 521
 RESULT 7
 POLG_HCVTW STANDARD; PRT; 3010 AA.
 ID POLG_HCVTW
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230206; PubMed=1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 "The Taiwanese hepatitis C virus genome: sequence determination and
 mapping the 5' termini of viral genomic and antigenomic RNA.";
 Virology 188:102-113(1992).
 RL
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 (RNA) (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M84754; -- NOT_ANNOTATED_CDS.
 DR PIR; A40244; GNWVTV.
 DR PDB; 1N64; 25-FEB-03.
 DR PDB; 1NS3; 08-APR-98.
 DR MEROPS; S29.001; --
 DR MEROPS; U39.001; --
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRP; 1.
 DR ProDom; PD186662; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CORE PROTEIN (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH_BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 55.0%; Score 11; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.0021;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVGTTDR 11
 |||||
 Db 511 PSPVVGTTDR 521

RESULT 8
 POLG_HCVJ6 STANDARD; PRT; 3033 AA.
 ID AC P26660;
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J6) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9204440; PubMed=1658196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 RA Machida A., Miyakawa Y., Mayumi M.;
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 RT from a human carrier: comparison with reported isolates for conserved
 RT and divergent regions.";
 RL J. Gen. Virol. 72:2697-2704(1991).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA}(N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D00944; BAA00792.1; -.
 DR PIR; JQ1303; JQ1303.
 DR HSSP; P27958; 1HE1.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.


```
FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 737 737
SQ SEQUENCE 737 AA; 81207 MW; 3AF69D82AD501B1 CRC64;

Query Match 50.0%; Score 10; DB 1; Length 737;
Best Local Similarity 100.0%; Pred. No. 0.0062; Mismatches 0; Gaps 0;
Matches 10; Conservative 0; Indels 0;

Qy 1 PSPVVVGTTD 10
Db 513 PSPVVVGTTD 522

RESULT 10
POLG_HCVJ8 STANDARD; PRT; 3033 AA.
AC P26651;
DT 01-AUG-1992 (Rel. 23, Created)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP58) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J8) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11115;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
RA Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D10988; BAA01761.1; -
CC PIR; A40250; GNWVJ8.
CC HSSP; P27958; 1HEI.
CC MEROPS; S23.001; -.
CC MEROPS; U39.001; -.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002522; HCV capsid.
CC InterPro; IPR002521; HCV_core.
```

```
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral RdRp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT TRANSMEM 347 369
FT ACT_SITE 1087 1087
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT NP_BIND 1324 1241
FT SITE 1320 1323
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 299 299
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT CARBOHYD 1091 1091
FT CARBOHYD 2038 2038
FT CARBOHYD 2359 2359
FT CARBOHYD 2811 2811
SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;

Query Match 50.0%; Score 10; DB 1; Length 3033;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTD 10
Db 513 PSPVVVGTTD 522
```



```

RESULT 11
ID POLG HCVH4 STANDARD; PRT; 520 AA.
AC Q01404;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (Fragment).
OS Hepatitis C virus (isolate HCV-476) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31643;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93019030; PubMed=1383400;
RA Abe K., Inchauspe G., Fujisawa K.;
RT "Genomic characterization and mutation rate of hepatitis C virus
RT isolated from a patient who contracted hepatitis during an epidemic
RT of non-A, non-B hepatitis in Japan.";
RL J. Gen. Virol. 73:2725-2729(1992).
CC -!- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; D10688; BAA01530.1; -
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
DR Transmembrane; Nonstructural protein.
KW INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 >520
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 418 418
FT CARBOHYD 424 424
FT CARBOHYD 431 431
FT CARBOHYD 449 449
FT NON_TER 520
SQ SEQUENCE 520 AA; 56499 MW; AA135246CF20D525 CRC64;

Query Match 45.0%; Score 9; DB 1; Length 520;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTG 9
Db 512 PSPVVVGTG 520

```

```

RESULT 12
ID POLG HCVHK STANDARD; PRT; 520 AA.
AC Q01403;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (Fragment).
OS Hepatitis C virus (isolate HCV-KP) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31644;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93019030; PubMed=1383400;
RA Abe K., Inchauspe G., Fujisawa K.;
RT "Genomic characterization and mutation rate of hepatitis C virus
RT isolated from a patient who contracted hepatitis during an epidemic
RT of non-A, non-B hepatitis in Japan.";
RL J. Gen. Virol. 73:2725-2729(1992).
CC -!- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; D10687; BAA01529.1; -
DR PIR; JQ1925; JQ1925.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
DR Transmembrane; Nonstructural protein.
KW INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 >520
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 418 418
FT CARBOHYD 424 424
FT CARBOHYD 431 431
FT CARBOHYD 449 449
FT NON_TER 520
SQ SEQUENCE 520 AA; 56476 MW; 1D2BD0A6FF27349B CRC64;

Query Match 45.0%; Score 9; DB 1; Length 520;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTG 9

```

```
Db          512 PSPVVGTT 520
|||||
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          1 PSPVW 6
          |||||
Db          35 PSPVV 40

RESULT 15
KRFT COLLI
ID KRFT COLLI STANDARD; PRT; 95 AA.
AC P07521;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Feather keratin B-4 (F-ker).
OS Columba livia (Domestic pigeon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Columbiformes; Columbidae; Columba.
OX NCBI_TaxID=8932;
RN [1]
RP SEQUENCE.
RC TISSUE=Feather;
RA Arai K.M., Takahashi R., Yokote Y., Akahane K.;
RT "The primary structure of feather keratins from duck (Anas platyrhynchos) and pigeon (Columba livia).";
RL Biochim. Biophys. Acta 873:6-12(1986).
CC -!- SUBUNIT: THE AVIAN KERATINS (F-KER, S-KER, C-KER AND B-KER) ARE A COMPLEX MIXTURE OF VERY SIMILAR POLYPEPTIDES.
CC -!- SIMILARITY: BELONGS TO THE AVIAN KERATIN FAMILY.
DR PIR; B24506; KRPF4.
DR InterPro: IPR003461; Keratin.
DR Pfam: PF02422; Keratin; 1.
KW Keratin; Feather; Multigene family; Acetylation.
FT MOD RES 1 1 ACETYLATION.
SQ SEQUENCE 95 AA; 9719 MW; 52E986B3329EASE8 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 95;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          1 PSPVW 6
          |||||
Db          35 PSPVV 40

Search completed: November 21, 2003, 20:59:57
Job time : 4.85 secs

Db          512 PSPVVGTT 520
|||||
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          3 PVVVGTT 8
          |||||
Db          44 PVVVGTT 49

RESULT 14
KRFT ANAPL
ID KRFT ANAPL STANDARD; PRT; 95 AA.
AC P08335;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Feather keratin B-4 (F-ker).
OS Anas platyrhynchos (Domestic duck).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anas.
OX NCBI_TaxID=8839;
RN [1]
RP SEQUENCE.
RC TISSUE=Feather;
RA Arai K.M., Takahashi R., Yokote Y., Akahane K.;
RT "The primary structure of feather keratins from duck (Anas platyrhynchos) and pigeon (Columba livia).";
RL Biochim. Biophys. Acta 873:6-12(1986).
CC -!- SUBUNIT: THE AVIAN KERATINS (F-KER, S-KER, C-KER AND B-KER) ARE A COMPLEX MIXTURE OF VERY SIMILAR POLYPEPTIDES.
CC -!- SIMILARITY: BELONGS TO THE AVIAN KERATIN FAMILY.
DR PIR; A24506; KRDF4.
DR InterPro: IPR003461; Keratin.
DR Pfam: PF02422; Keratin; 1.
KW Keratin; Feather; Multigene family; Acetylation.
FT MOD RES 1 1 ACETYLATION.
SQ SEQUENCE 95 AA; 9662 MW; B90FC4CA43805817 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 95;
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:45:57 ; Search time 24.4 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-81

Perfect score: 20
Sequence: 1 PSPVVVGTTDRSGAPTYSWG 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	187	Q9PXG5	Q9pxg5 hepatitis c
2	20	100.0	206	Q81571	Q81571 hepatitis c
3	20	100.0	283	Q68446	Q68446 hepatitis c
4	20	100.0	283	Q68441	Q68441 hepatitis c
5	20	100.0	283	Q68442	Q68442 hepatitis c
6	20	100.0	283	Q68445	Q68445 hepatitis c
7	20	100.0	490	Q9PXN2	Q9pxn2 hepatitis c
8	20	100.0	778	Q84184	Q84184 hepatitis c
9	20	100.0	2436	Q81756	Q81756 hepatitis c
10	20	100.0	2864	Q92975	Q92975 hepatitis c
11	20	100.0	3010	Q90193	Q90193 hepatitis c
12	20	100.0	3010	Q68833	Q68833 hepatitis c
13	20	100.0	3010	Q90194	Q90194 hepatitis c
14	20	100.0	3010	Q9DTE8	Q9dte8 hepatitis c
15	20	100.0	3011	Q9IFES	Q9ife5 hepatitis c
16	20	100.0	3011	Q36579	Q36579 hepatitis c

17	20	100.0	3011	12	O36610	O36610 hepatitis c
18	20	100.0	3011	12	Q9ELS8	Q9eles hepatitis c
19	20	100.0	3011	12	O36609	O36609 hepatitis c
20	20	100.0	3011	12	O36608	O36608 hepatitis c
21	18	90.0	364	12	O36980	O36980 hepatitis c
22	17	85.0	137	12	Q914Q6	Q914q6 hepatitis c
23	17	85.0	137	12	Q914Q5	Q914q5 hepatitis c
24	17	85.0	137	12	O914Q8	O914q8 hepatitis c
25	17	85.0	137	12	O914Q4	O914q4 hepatitis c
26	17	85.0	137	12	O914W0	O914w0 hepatitis c
27	17	85.0	137	12	O914Q0	O914q0 hepatitis c
28	17	85.0	137	12	Q914V9	Q914v9 hepatitis c
29	17	85.0	137	12	Q914Q3	Q914q3 hepatitis c
30	17	85.0	137	12	O914Q1	O914q1 hepatitis c
31	17	85.0	137	12	O914Q7	O914q7 hepatitis c
32	17	85.0	137	12	O914V7	O914v7 hepatitis c
33	17	85.0	189	12	O9YZ08	O9yzq8 hepatitis c
34	17	85.0	420	12	Q98UN4	Q98un4 hepatitis c
35	17	85.0	640	12	Q68966	Q68966 hepatitis c
36	17	85.0	778	12	O04185	O04185 hepatitis c
37	17	85.0	782	12	O68951	O68951 hepatitis c
38	17	85.0	1008	12	O89153	O89153 hepatitis c
39	17	85.0	1008	12	O89154	O89154 hepatitis c
40	17	85.0	1008	12	O89152	O89152 hepatitis c
41	17	85.0	3010	12	O9DTE2	O9dte2 hepatitis c
42	17	85.0	3010	12	O9J3G2	O9j3g2 hepatitis c
43	17	85.0	3010	12	P88803	P88803 hepatitis c
44	17	85.0	3010	12	O9J3H8	O9j3h8 hepatitis c
45	17	85.0	3010	12	O9J3G5	O9j3g5 hepatitis c

ALIGNMENTS

RESULT 1 0
Q9PXG5 PRELIMINARY; PRT; 187 AA.
ID O9PXG5;
AC 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Genome polyprotein [Contains: envelope glycoprotein E2/NS1 (GP68)]
DE (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95197817; PubMed=7534322;
RA Zonaro A., Ravaggi A., Puoti M., Krensdorf D., Albertini A.,
RA Cariani E.;
RT "Differential pattern of sequence heterogeneity in the hepatitis C
virus E1 and E2/NS1 proteins.";
RL J. Hepatol. 21:858-865(1994).
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 187 AA; 20470 MW; 30DE4B45566A5602 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 187;
Best Local Similarity 100.0%; Pred. No. 7.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PSPVVVGTTDRSGAPTYSWG 20
|||||

Db 153 PSPVVVGTTDRSGAPTYSWG 172
|||||

```

RESULT 2
Q81571 ID Q81571 PRELIMINARY; PRT; 206 AA.
AC Q81571;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE NS2 protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=Hutchinson;
RX MEDLINE=91013116; PubMed=2170712;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA Yoshizawa H.;
RT "The 5'-terminal sequence of the hepatitis C virus genome.";
RL Jpn. J. Exp. Med. 60:167-177(1990).
RN [2]_
RP SEQUENCE FROM N.A.
RC STRAIN=Hutchinson;
RA Inchausti G., Zebedee S.L., Nasoff M.S., Sugitani M., Abe K.,
RA Prince A.M.;
RT "Cloning and nucleotide sequence analysis of structural and
RT nonstructural regions of the hutchinson strain of hepatitis C.";
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
RN [3]_
RP SEQUENCE FROM N.A.
RC STRAIN=Hutchinson;
RX MEDLINE=89222455; PubMed=2496467;
RA Kuo G., Choo Q.-L., Alter H.J., Gitnick G.L., Redeker A.G.,
RA Purcell R.H., Miyamura T., Dienstag J.L., Alter M.J., Stevens C.E.,
RA Tegmeyer G.E., Bonino F., Colombo M., Lee W.-S., Kuo C., Berger K.,
RA Shuster J.R., Overby L.R., Bradley D.W., Houghton M.;
RT "An assay for circulating antibodies to a major etiologic virus of
RT human non-A, non-B hepatitis.";
RL Science 244:362-365(1992).
DR EMBL; M55971; AAA45615.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR CoRoT; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 206
SQ SEQUENCE 206 AA; 22807 MW; 7FC6174D92325295 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 206;
Best Local Similarity 100.0%; Pred. No. 7.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDRSGAPTYSWG 20
Db 21 PSPVVVGTTDRSGAPTYSWG 40

RESULT 3
Q68446 ID Q68446 PRELIMINARY; PRT; 283 AA.
AC Q68446;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]_
RP SEQUENCE FROM N.A.
RA Vizmanos J.L.;

```

```

RT "Analysis of the 5'NC, E1, E2 and NS5 genome regions of the hepatitis
RT C virus.";
RL Thesis (1996), Genetics Dept., University of Navarra.
DR EMBL; U37623; AAA86567.1; -.
DR InterPro; IPR002519; HCV NS1.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV NS1; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 283
SQ SEQUENCE 283 AA; 30635 MW; F11C115A282E5FFF CRC64;

Query Match 100.0%; Score 20; DB 12; Length 283;
Best Local Similarity 100.0%; Pred. No. 1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDRSGAPTYSWG 20
Db 249 PSPVVVGTTDRSGAPTYSWG 268

RESULT 4
Q68441 ID Q68441 PRELIMINARY; PRT; 283 AA.
AC Q68441;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]_
RP SEQUENCE FROM N.A.
RA Vizmanos J.L.;
RT "Analysis of the 5'NC, E1, E2 and NS5 genome regions of the hepatitis
RT C virus.";
RL Thesis (1996), Genetics Dept., University of Navarra.
DR EMBL; U37618; AAA86562.1; -.
DR InterPro; IPR002519; HCV NS1.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV NS1; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 283
SQ SEQUENCE 283 AA; 30601 MW; 415CBB5A228EF232 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 283;
Best Local Similarity 100.0%; Pred. No. 1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSPVVVGTTDRSGAPTYSWG 20
Db 249 PSPVVVGTTDRSGAPTYSWG 268

RESULT 5
Q68442 ID Q68442 PRELIMINARY; PRT; 283 AA.
AC Q68442;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

```

```
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RA Vizmanos J.L.;
RT "Analysis of the 5'NC, E1, E2 and NS5 genome regions of the hepatitis C virus.";
RL Thesis (1996), Genetics Dept., University of Navarra.
DR EMBL; U37619; AAA86563.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 283
SQ SEQUENCE 283 AA; 30601 MW; 415CBB5A228EF232 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 283;
Best Local Similarity 100.0%; Pred. No. 1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTDRSGAPTYSWG 20
Db 249 PSPVVVGTDRSGAPTYSWG 268

RESULT 6
Q68445
ID Q68445 PRELIMINARY; PRT; 283 AA.
AC Q68445;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DE 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Vizmanos J.L.;
RT "Analysis of the 5'NC, E1, E2 and NS5 genome regions of the hepatitis C virus.";
RL Thesis (1996), Genetics Dept., University of Navarra.
DR EMBL; U37622; AAA86566.1; -.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 283
SQ SEQUENCE 283 AA; 30601 MW; 415CBB5A228EF232 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 283;
Best Local Similarity 100.0%; Pred. No. 1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTDRSGAPTYSWG 20
Db 249 PSPVVVGTDRSGAPTYSWG 268

RESULT 7
Q9PXN2
ID Q9PXN2 PRELIMINARY; PRT; 490 AA.
AC Q9PXN2;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
```

```
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Genome polyprotein [Contains: envelope glycoprotein E2/NS1 (GP68)] (Fragment).
OS Hepatitis C virus type 2.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=40271;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94351179; PubMed=7520922;
RA Nakazawa T., Kato N., Ohkoshi S., Shibuya A., Shimotohno K.;
RL J. Hepatol. 20:623-629 (1994).
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 490 AA; 53572 MW; 2C61F302F307D63C CRC64;

Query Match 100.0%; Score 20; DB 12; Length 490;
Best Local Similarity 100.0%; Pred. No. 1.7e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTDRSGAPTYSWG 20
Db 350 PSPVVVGTDRSGAPTYSWG 369

RESULT 8
Q04184
ID Q04184 PRELIMINARY; PRT; 778 AA.
AC Q04184; Q81810;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91195357; PubMed=1849654;
RA Ogata N., Alter H.J., Miller R.H., Purcell R.H.;
RT "Nucleotide sequence and mutation rate of the H strain of hepatitis C virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:3392-3396 (1991).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; M62381; AAB02127.1; -.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT CHAIN 514 >778 STRUCTURAL PROTEIN.
FT NON_TER 778
SQ SEQUENCE 778 AA; 85143 MW; F7709172CD03E39B CRC64;
```

DT	01-NOV-1998	(TrEMBLrel. 08, Last sequence update)
DE	01-MAR-2003	(TrEMBLrel. 23, Last annotation update)
DE	Genome polyprotein (Fragment).	
OS	Hepatitis C virus.	
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;	
OC	Hepadnavirus.	
OX	NCHI_TaxID=11103;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=HC-J4;	
RX	MEDLINE=98240944; PubMed=9581788;	
RA	Yanagih J., St Clare M., Shapiro M., Emerson S.U., Purcell R.H.,	
RA	Bukh J.;	
RT	"Transcripts of a chimeric clone of hepatitis C virus genotype 1b are	
RT	transfected in vivo,"	
RL	Virology 244:161-172(1998).	
CC	-I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A	
CC	LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:	
CC	PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF	
CC	PROTEIN C AND MRNA (BY SIMILARITY).	
EMBL	AF054258; AAC15731.1; --	
HSSP	P26663; 1JXP	
DR	InterPro; IPR001410; DEAD.	
DR	InterPro; IPR002521; HCV_capsid.	
DR	InterPro; IPR002521; HCV_core.	
DR	InterPro; IPR002519; HCV env.	
DR	InterPro; IPR002531; HCV NS1.	
DR	InterPro; IPR002518; HCV NS2.	
DR	InterPro; IPR004109; HCV NS3.	
DR	InterPro; IPR000745; HCV NS4a.	
DR	InterPro; IPR001490; HCV NS4b.	
DR	InterPro; IPR002868; HCV NS5a.	
DR	InterPro; IPR002166; HCV RdRp.	
DR	InterPro; IPR007095; RNA_pol_DS_PS.	
DR	InterPro; IPR007094; RNA_pol_PSVir.	
DR	Pfam; PF01543; HCV capsid; 1.	
DR	Pfam; PF01542; HCV core; 1.	
DR	Pfam; PF01539; HCV env; 1.	
DR	Pfam; PF01560; HCV NS1; 1.	
DR	Pfam; PF01538; HCV NS2; 1.	
DR	Pfam; PF02907; HCV NS3; 1.	
DR	Pfam; PF01006; HCV NS4a; 1.	
DR	Pfam; PF01001; HCV NS4b; 1.	
DR	Pfam; PF01506; HCV NS5a; 1.	
DR	Pfam; PF00998; Viral_RdRp; 1.	
DR	ProDom; PD186062; HCV_NS1; 1.	
DR	SMART; SM00487; DEXDC; 1.	
DR	PROSITE; PS50507; RDRF_POSITIVE; 1.	
DR	PROSITE; PS50521; RDRF_VIRAL; 1.	
KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;	
KW	Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.	
FT	NON_TER 2864 2864	
SQ	SEQUENCE 2864 AA; 310415 MW; COCD3933ED07C6A5 CRC64;	
Query Match 100.0%; Score 20; DB 12; Length 2864;		
Best Local Similarity 100.0%; Pred.No. 8.1e-12;		
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Oy	1 PSPVVVGTTDRSGAPTYSWG 20	
Db	511 PSPVVVGTTDRSGAPTYSWG 530	
RESULT 11		
P90193	ID P90193 PRELIMINARY; PERT; 3010 AA.	
AC	P90193;	
DT	01-MAY-1997 (TrEMBLrel. 03, Created)	
DT	01-MAY-1997 (TrEMBLrel. 03, Last sequence update)	
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)	
DE	Genome polyprotein.	
OS	Hepatitis C virus.	
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;	

```

OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-1b;
RA Enomoto N.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-1b;
RX MEDLINE=95340824; PubMed=7542279;
RA Enomoto N., Sakuma I., Asahina Y., Kurosaki M., Murakami T.,
RA Yamamoto C., Izumi N., Marumo F., Sato C.;
RT "Comparison of full-length sequences of interferon-sensitive and
RT resistant hepatitis C virus 1b."
RL J. Clin. Invest. 96:224-230(1995).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; D50485; BAA09076.1; -.
DR HSP; P26663; 1JXP.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA pol DS_PS.
DR InterPro; IPR007094; RNA pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; helicase C; 1.
DR Pfam; PF00998; Viral RdRp; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50321; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
FT CHAIN 1 191
FT CHAIN 192 383
FT CHAIN 384 809
FT CHAIN 810 1026
FT CHAIN 1027 1657
FT CHAIN 1658 1711
FT CHAIN 1712 1972
FT CHAIN 1973 2419
FT CHAIN 2420 3010
FT CHAIN 3010 3010
SQ SEQUENCE 3010 AA; 326885 MW; 21CD35B3DAC02B84 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 3010;
Best Local Similarity 100.0%; Pred. No. 8.4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTDRSGAPTYSWG 20
DB 511 PSPVVVGTTDRSGAPTYSWG 530

RESULT 13
P90194
ID P90194 PRELIMINARY; PRT; 3010 AA.

```

P90194;
 AC 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Hepatitis C virus.
 OC Hepacivirus.
 OC NCBI_TaxID=11103;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=HCV-1b;
 RC MEDLINE=95340824; PubMed=7542279;
 RX Enomoto N., Sakuma I., Asahina Y., Kurosaki M., Murakami T.,
 RA Yamamoto C., Izumi N., Marumo F., Sato C.;
 RT "Comparison of full-length sequences of interferon-sensitive and
 RT resistant hepatitis C virus 1b."
 RL J. Clin. Invest. 96:224-230(1995).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MNA (BY SIMILARITY).
 DR EMBL; D50481; BAA09072.1; -.
 DR HSP; P26663; LUXP.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRP.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR007095; RNA pol DS PS.
 DR InterPro; IPR007094; RNA pol PSvir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS050507; RDRP_POSITIVE; 1.
 DR PROSITE; PS50521; RDRP_VIRAL; 1.
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydrolase; Nonstructural protein; Polyprotein;
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 FT CHAIN 1 191
 FT CHAIN 192 383
 FT CHAIN 384 809
 FT CHAIN 810 1026
 FT CHAIN 1027 1657
 FT CHAIN 1658 1711
 FT CHAIN 1712 1972
 FT CHAIN 1973 2419
 FT CHAIN 2420 3010
 SQ SEQUENCE 3010 AA; 326914 MW; 98D5C2A2D47FD011 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 3010;

Best Local Similarity 100.0%; Pred. No. 8.4e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 PSPVVVGTTDRSGAPTYSWG 20
 Db 511 PSPVVVGTTDRSGAPTYSWG 530
 RESULT 14
 Q9DTE8
 ID Q9DTE8 PRELIMINARY; PRT; 3010 AA.
 AC Q9DTE8;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HCV1109;
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
 RA Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
 RA Mishiro S.;
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
 RT with hepatocellular carcinoma: the 'progression score' revisited."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MNA (BY SIMILARITY).
 DR EMBL; AB049089; BAB18802.1; -.
 DR HSP; P26663; LUXP.
 DR InterPro; IPR000345; CytC_heme_bind.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRP.
 DR InterPro; IPR007095; RNA pol DS PS.
 DR InterPro; IPR007094; RNA pol PSvir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS050507; RDRP_POSITIVE; 1.
 DR PROSITE; PS50521; RDRP_VIRAL; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327067 MW; E26F4D6569A836C80 CRC64;
 Query Match 100.0%; Score 20; DB 12; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 8.4e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 PSPVVVGTTDRSGAPTYSWG 20

Search completed: November 21, 2003, 21:08:18
Job time : 24.45 secs

```
Db 511 PSPVVVGTTRDSCGPTYSWG 530
|||||
RESULT 15
Q9IFES PRELIMINARY; PRT; 3011 AA.
AC Q9IFES;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RX MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
RT sequence."
RL J. Gen. Virol. 82:1291-1297(2001).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF271632; AAF81759.1;
DR HSSP; P27958; 1A1V.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NSI.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002186; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR SMART; SM00487; DEXDc; 1.
DR PROSITE; PS00150; CYTOCHROME_C; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327124 MW; 2489CE74AC864E58 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 8.4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSPVVVGTTRDSCGPTYSWG 20
|||||
Db 511 PSPVVVGTTRDSCGPTYSWG 530
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:22 ; Search time 32.1 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-82

Perfect score: 20

Sequence: 1 GAPTYSWGENDTVFLNLT 20

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq 19Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	16	AA1984511
2	20	100.0	20	17	AA1991005
3	20	100.0	20	23	AA1981711
4	20	100.0	179	21	AA198529
5	20	100.0	254	22	AA198043
6	20	100.0	363	22	AA198042
7	20	100.0	402	14	AA198438
8	20	100.0	454	10	AA1990183
9	20	100.0	454	21	AA198526

10	20	100.0	463	14	AA1933588	HCV CKS-NS1S1 fusi
11	20	100.0	463	14	AA1933187	Sequence of subfra
12	20	100.0	463	22	AA1969007	HCV recombinant an
13	20	100.0	480	14	AA1933992	HCV-1 E2/NS1 prote
14	20	100.0	531	22	AA1902622	Chimeric HCV E2/661
15	20	100.0	621	14	AA1933185	Sequence of subfra
16	20	100.0	622	14	AA1933591	HCV CKS-NS1S1-NS1S
17	20	100.0	622	22	AA1969010	HCV recombinant an
18	20	100.0	637	24	AA197410	Hepatitis C virus
19	20	100.0	663	17	AA192935	HCV E2 + NS2 poly
20	20	100.0	663	20	AA197615	Hepatitis C virus
21	20	100.0	738	14	AA1933592	HCV CKS-full lengt
22	20	100.0	738	22	AA1969011	HCV recombinant an
23	20	100.0	2435	13	AA1925135	HCV polypeptide 1.
24	20	100.0	2436	10	AA192050	Sequence encoded i
25	20	100.0	2436	10	AA190288	Peptide encoded by
26	20	100.0	2436	13	AA1928582	HCV amino acid seq
27	20	100.0	2772	11	AA1908123	Hepatitis C virus
28	20	100.0	2772	21	AA1918540	Protein encoded by
29	20	100.0	2816	14	AA1934009	HCV-1 polyprotein.
30	20	100.0	2894	13	AA1924440	Composite HCV HC-J
31	20	100.0	2894	16	AA1970230	Composite hepatitis
32	20	100.0	2955	11	AA1908124	Hepatitis C virus
33	20	100.0	2955	20	AA1914975	Amino acid sequenc
34	20	100.0	2955	21	AA1918541	Polyprotein encode
35	20	100.0	3011	13	AA1921519	Compiled HCV sequ
36	20	100.0	3011	14	AA1931621	Hepatitis C virus
37	20	100.0	3011	17	AA1909031	Hepatitis C virus
38	20	100.0	3011	18	AA1934480	HCV polyprotein.
39	20	100.0	3011	19	AA1940038	HCV polyprotein.
40	20	100.0	3011	23	AA1922049	Hepatitis C virus
41	20	100.0	3011	23	AA1922052	Hepatitis C virus
42	13	65.0	192	19	AA1967009	HCV nucleocapsid c
43	11	55.0	30	23	AA1984633	HCV HepC1a segment
44	11	55.0	36	21	AA1989700	Core polypeptide f
45	11	55.0	36	22	AA1901108	Viral DPI78/107-11

ALIGNMENTS

RESULT 1
AA1984511
ID AA1984511 standard; peptide; 20 AA.
AC AA1984511;
XX
XX
DT 06-JAN-1997 (first entry)
XX
XX
DE Hepatitis C virus peptide NS1-13 (residues 523-542).
XX
XX
KW Hepatitis C virus; HCV; immunogen; non-structural region; NS1;
KW immunodominant; T cell epitope; vaccine.
XX
OS Hepatitis C virus.
XX
XX
PW WO9512677-A2.
XX
XX
PD 11-MAY-1995.
XX
XX
PF 28-OCT-1994; 94WO-EP03555.
XX
XX
PR 04-NOV-1993; 93EP-0402718.
XX
XX
PA (INNO-) INNOGENETICS NV.
XX
XX
PI Deleys R, Leroux-Roels G, Maertens G;
XX
XX
DR WPI; 1995-193822/25.
XX
XX
PT Hepatitis C Virus immunogenic polypeptide contg. a T-cell
PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in
PT production of vaccines, therapeutic agents, etc.

XX Example 4; Page 51; 105pp; English.

PS A series of overlapping peptides (including the present sequence) was

CC synthesised based on sequences in the core, E1 and E2/NS1 regions of

CC hepatitis C virus. The peptides were used as antigens in lympho-

CC proliferative assays to identify the main T-cell epitopes.

XX Sequence 20 AA;

SQ Query Match 100.0%; Score 20; DB 16; Length 20;

Best Local Similarity 100.0%; Pred. No. 2.3e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAPTYSWGENDTDVFLNNT 20

DB 1 GAPTYSWGENDTDVFLNNT 20

RESULT 2

AAR91005

ID AAR91005 standard; peptide; 20 AA.

XX AC AAR91005;

XX DT 25-SEP-1996 (first entry)

XX DE HCV E2 peptide E2-13B for competition studies.

XX KW HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;

XX KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.

XX OS Synthetic.

XX PN WO9604385-A2.

XX PD 15-FEB-1996.

XX PF 31-JUL-1995; 95WO-EP03031.

XX PR 29-JUL-1994; 94EP-0870132.

XX PA (INNO-) INNOGENETICS NV.

XX PI Bosman F, Buyse M, De Martynoff G, Maertens G;

XX DR WPI; 1996-129401/13.

XX PT Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope

PT proteins - in presence of disulphide bond cleavage agent, to

PT produce proteins suitable for direct use in vaccines or diagnostic

PT assays of HCV

XX Claim 29; Page 67; 146pp; English.

XX AAR90965-R90980 and AAR90995-R91015 represent synthetic hepatitis C

CC virus (HCV) E1 and E2 peptides used in competition studies. This

CC sequence represents a synthetic E2 peptide, and corresponds to residues

CC 523-542 of the E2 protein sequence. These sequences are useful for in

CC vitro monitoring of HCV disease, or prognosis of the response to

CC interferon treatment of patients suffering from HCV infection. These

CC sequences compete with the proteins produced by AAT12704-T12709 and

CC AAT12961-T12974, which are included in vectors for the production of

CC recombinant E1, E2, and E1/E2. The recombinant proteins can then be

CC isolated and purified by carrying out a disulphide bond cleavage, or a

CC reduction step with a disulphide bond cleavage agent, after lysis of

CC recombinant host cells. The constructs containing the purified HCV

CC envelope proteins can be used for vaccinating humans against HCV, for in

CC vitro detection of HCV antibodies in a sample, and in a serotyping assay

CC for detecting one or more serological types of HCV present in a

CC biological sample. The constructs can also be immobilised on a solid

CC substrate and incorporated into a reversed phase hybridisation assay for

CC determining the presence or the genotype of HCV. The new purification

CC method preserves the conformation of the recombinantly expressed E1, E2

CC and E1/E2, and eliminates contaminating proteins. Antigens isolated

CC using this method are more reactive with human sera than those isolated

CC by known techniques.

XX Sequence 20 AA;

SQ Query Match 100.0%; Score 20; DB 17; Length 20;

Best Local Similarity 100.0%; Pred. No. 2.3e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAPTYSWGENDTDVFLNNT 20

DB 1 GAPTYSWGENDTDVFLNNT 20

RESULT 3

AAO18711

ID AAO18711 standard; Peptide; 20 AA.

XX AC AAO18711;

XX DT 24-OCT-2002 (first entry)

XX DE Hepatitis C virus E2 protein derived peptide E2-13B.

XX KW Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide;

XX KW immunostimulant; vaccine.

XX OS Hepatitis C virus.

XX PN WO200255548-A2,

XX PD 18-JUL-2002.

XX PF 11-JAN-2002; 2002WO-EP00219.

XX PR 11-JAN-2001; 2001US-260699P.

XX PR 30-AUG-2001; 2001US-315768P.

XX PA (INNO-) INNOGENETICS NV.

XX PI Maertens G, Bosman F, Buyse M;

XX DR WPI; 2002-599657/64.

XX PT New therapeutic vaccine compositions comprising at least one purified

PT recombinant hepatitis C virus (HCV) single or specific oligomeric

PT recombinant envelope protein E1 or E2, useful for immunizing humans

XX from HCV infection

XX Claim 4; Page 228; 243pp; English.

XX The present invention relates to new therapeutic vaccine compositions for

CC inducing hepatitis C virus (HCV)-specific antibodies, comprising a

CC composition containing at least one purified recombinant HCV single or

CC specific oligomeric recombinant envelope proteins selected from an E1 and

CC an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are

CC useful for inducing HCV-specific antibodies or for immunising humans

CC against HCV. The recombinant HCV E1 and/or E2 proteins are useful as

CC vaccines or therapeutics, in HCV screening and confirmatory antibody

CC tests, for raising antibodies, in the preparation of medicament, and for

CC in vitro monitoring of HCV disease or prognosing the response to

CC treatment of patients suffering from HCV infection. The present sequence

CC is a peptide derived from the proteins of the invention.

XX Sequence 20 AA;

SQ Query Match 100.0%; Score 20; DB 23; Length 20;

Best Local Similarity 100.0%; Pred. No. 2.3e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAPTYSWGENDTDVFLNNT 20

DB 1 GAPTYSWGENDTDFVLNNT 20
|||||

RESULT 4

AAB18529
ID AAB18529 standard; Protein; 179 AA.

AC AAB18529;

DT 15-JAN-2001 (first entry)

DE Protein encoded by a novel hepatitis C virus cDNA clone 13i.

XX Hepatitis C virus; HCV; antisense polynucleotide; polypeptide;

KW viral infectivity; viral replication.

XX Hepatitis C virus.

OS EP1034785-A2.

PN 13-SEP-2000.

XX 16-MAR-1990; 2000EP-0109602.

PR 17-MAR-1989; 89US-0325338.

PR 20-APR-1989; 89US-0341334.

PR 18-MAY-1989; 89US-0355002.

PR 16-MAR-1990; 90EP-0302866.

XX (CHIR) CHIRON CORP.

PA Houghton M, Choo Q, Kuo G;

PI WPI; 2000-566891/53.

DR N-PSDB; AAA75285.

XX Novel composition comprising a hepatitis C virus antisense polynucleotide which is complementary to or corresponds to a sense strand of the virus genome, and selectively hybridizes to it -

XX Example; Fig 5; 75pp; English.

CC The specification describes a pharmaceutical composition which comprises a hepatitis C virus (HCV) antisense polynucleotide. The HCV is characterized by a positive stranded RNA genome which has 40% homology at the polypeptide level to a HCV polypeptide. The antisense polynucleotide binds to cellular polynucleotides which enhance and/or are required for viral infectivity, replicative ability or chronicity. The antisense polynucleotides may also be designed to bind with high specificity, to be of increased stability, to be stable and to have low toxicity. The composition also comprises an agent which causes viral RNA to be inactive. The composition is used for preventing HCV replication in a system. The present sequence is encoded by a novel HCV cDNA sequence, which is used in the course of the invention.

XX Sequence 179 AA;

Query Match 100.0%; Score 20; DB 21; Length 179;

Best Local Similarity 100.0%; Pred. No. 1.6e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDFVLNNT 20

DB 13 GAPTYSWGENDTDFVLNNT 32

RESULT 5

AAB68043
ID AAB68043 standard; protein; 254 AA.

XX AAB68043;

XX DT 29-JUN-2001 (first entry)

XX Amino acid sequence of water soluble variant of envelope E2 protein.

XX Eo protein; HCV; envelope protein; E2 protein; HCV infection;

KW HCV attachment.

XX Synthetic.

OS Hepatitis C virus.

XX WO200122984-A1.

PN 05-APR-2001.

XX 26-SEP-2000; 2000WO-US26395.

XX 29-SEP-1999; 99US-0407430.

XX (UYCO) UNIV COLUMBIA NEW YORK.

XX Worman HJ, Mamiya N;

XX WPI; 2001-273486/28.

XX Treating or preventing hepatitis C virus infection in a subject, involves administering hepatitis C virus envelope protein E2 binding agents -

PS Claim 5; Fig 8; 46pp; English.

XX The present sequence represents a water soluble variant of a Hepatitis C virus (HCV) envelope E2 protein. Agents that bind to the HCV E2 protein (such as the human Eo protein), and so inhibit the attachment of HCV onto cells (especially liver cells), are used to treat HCV infections in mammals, in particular humans. The specification also describes a method for identifying a compound which can be used for treating or preventing HCV in a subject and which can inhibit the attachment of HCV onto cells by inhibiting the binding of HCV envelope E2 protein to a cellular protein associated with HCV attachment and entry into cells. The method comprises incubating the compound, HCV envelope E2 protein or its variant and a cellular protein capable of specifically binding to the HCV E2 protein under suitable reaction conditions; determining the interactions between HCV envelope E2 protein and cellular protein in the presence and absence of the compound; and comparing the interaction to identify a compound which can inhibit the attachment of HCV onto cells.

XX Sequence 254 AA;

Query Match 100.0%; Score 20; DB 22; Length 254;

Best Local Similarity 100.0%; Pred. No. 2.2e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDFVLNNT 20

DB 118 GAPTYSWGENDTDFVLNNT 137

RESULT 6

AAB68042

ID AAB68042 standard; protein; 363 AA.

XX AAB68042;

XX 29-JUN-2001 (first entry)

XX Amino acid sequence of a Hepatitis C virus envelope E2 protein.

XX Eo protein; HCV; envelope protein; E2 protein; HCV infection;

KW HCV attachment.

XX Hepatitis C virus.

XX

PS Disclosure; fig 46; 235pp; English.

XX The sequence is the peptide encoded by the hepatitis C virus
 CC (HCV) cDNA insert in clone k9-1 (see AAN90335). The polypeptides
 CC are used to diagnose HCV-induced NANBH, to raise antibodies for
 CC immunoassay or treatment, or to produce vaccines.
 CC The region shown overlaps the cDNA of AAN90327.
 CC (Updated on 25-MAR-2003 to correct PR field.)
 XX

SQ Sequence 454 AA;

Query Match 100.0%; Score 20; DB 10; Length 454;
 Best Local Similarity 100.0%; Pred. No. 3.7e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDDVFLNNT 20
 |||||
 Db 73 GAPTYSWGENDDVFLNNT 92

RESULT 9

ID AAB18526
 XX AAB18526 standard; Protein; 454 AA.

AC AAB18526;

XX 15-JAN-2001 (first entry)

XX Protein encoded by a novel hepatitis C virus cDNA clone k9-1.

XX Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
 KW viral infectivity; viral replication.

XX Hepatitis C virus.

XX EP1034785-A2.

XX 13-SEP-2000.

XX 16-MAR-1990; 2000EP-0109602.

XX 17-MAR-1989; 89US-0325338.

XX 20-APR-1989; 89US-0341334.

XX 18-MAY-1989; 89US-0355002.

XX 16-MAR-1990; 90EP-0302866.

XX (CHIR) CHIRON CORP.

XX Houghton M, Choo Q, Kuo G;

XX WPI; 2000-566891/53.

XX N-PSDB; AAA75282.

XX Novel composition comprising a hepatitis C virus antisense
 PT polynucleotide which is complementary to or corresponds to a sense
 PT strand of the virus genome, and selectively hybridizes to it -
 XX

PS Example; Fig 2; 75pp; English.

XX The specification describes a pharmaceutical composition which
 CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
 CC HCV is characterized by a positive stranded RNA genome which has
 CC 40% homology at the polypeptide level to a HCV polyprotein. The
 CC antisense polynucleotide binds to cellular polynucleotides which
 CC enhance and/or are required for viral infectivity, replicative
 CC ability or chronicity. The antisense polynucleotides may also be
 CC designed to bind with high specificity, to be of increased stability,
 CC to be stable and to have low toxicity. The composition also comprises
 CC an agent which causes viral RNA to be inactive. The composition
 CC is used for preventing HCV replication in a system. The present
 CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
 CC course of the invention.

XX

SQ Sequence 454 AA;

Query Match 100.0%; Score 20; DB 21; Length 454;
 Best Local Similarity 100.0%; Pred. No. 3.7e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDDVFLNNT 20
 |||||
 Db 73 GAPTYSWGENDDVFLNNT 92

RESULT 10

AAR33588
 ID AAR33588 standard; protein; 463 AA.

XX AAR33588;

XX 25-MAR-2003 (updated)

DT 05-JUL-1993 (first entry)

XX HCV CKS-NS1S1 fusion antigen.

XX Hepatitis C virus; non-A, non-B hepatitis virus; NANBH;

XX non-structural protein; CMP-KDO synthetase; CKS fusion protein;

XX CTP: CMP-3-deoxy-manno-octulosonate cytidyl transferase;

XX immunoassay; pHCV-77.

XX Hepatitis C virus.

XX WO9304088-A1.

XX 04-MAR-1993.

XX 21-AUG-1992; 92WO-US07188.

XX 21-AUG-1991; 91US-0748561.

XX (ABBO) ABBOTT LAB.

XX Dailey SH, Desai SM, Devare SG;

XX WPI; 1993-093941/11.

XX Hepatitis C assay using recombinant NS1 region antigens - for
 PT detecting antibodies and antigen in body fluids from individuals
 PT exposed to hepatitis C virus

XX Claim 1; Page 38-40; 175pp; English.

XX Eight oligonucleotides representing amino acids 365-579 of the HCV
 CC genome were ligated together and cloned as a 645bp EcoRI/BamHI
 CC fragment into the CKS fusion vector pJO200. The amino acid sequence
 CC of this antigen is designated pHCV-77 (i.e. AAR33588). The resultant
 CC fusion protein HCV CKS-NS1S1 consists of 239 amino acids of CKS,
 CC seven amino acids contributed by linker DNA sequences and 215 amino
 CC acids from the NS1 region of the HCV genome. The fusion protein is
 CC used to detect antibodies and antigen in body fluids from
 CC individuals exposed to HCV.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 463 AA;

Query Match 100.0%; Score 20; DB 14; Length 463;
 Best Local Similarity 100.0%; Pred. No. 3.8e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDDVFLNNT 20
 |||||
 Db 407 GAPTYSWGENDDVFLNNT 426

RESULT 11

AAR33187

```

ID AAR33187 standard; peptide; 463 AA.
XX
AC AAR33187;
XX
AC AAR33187;
XX
DT 25-MAR-2003 (updated)
DT 03-JUL-1993 (first entry)
XX
DE Sequence of subfragment pHCV65 (AAs 565-731) of the hepatitis
DE C virus (HCV) genome.
XX
KW Immunogenic peptide; hepatitis C virus; immunogenic domain;
KW monoclonal antibody; diagnosis; detection; therapy.
XX
OS Synthetic.
XX
FN WO9304205-A1.
XX
PD 04-MAR-1993.
XX
PF 21-AUG-1992; 92WO-US07189.
XX
PR 21-AUG-1991; 91US-0748292.
XX
PA (ABBO ) ABBOTT LAB.
XX
PI Dailey SH, Desai SM, Devare SG, Johnson JE, Mehta SU;
XX WPI; 1993-094030/11.
XX
PT Monoclonal antibodies specific for hepatitis C virus E2-NS1
PT antigen - useful for diagnosis and evaluation of HCV infections
PT and in differentiation studies
XX
PS Example; Pages 36-37; 48pp; English.
XX
CC Immunogenic domains of E2/NS1 region of HCV genome encompassing AAs
CC 600-720 were mapped with PRPSCAN analysis. Based on the EIA reactivity
CC of a panel of HCV positive sera, peptide AAR33184 was chosen as the
CC immunogen for the generation of monoclonal antibodies to HCV NS1.
CC Several individual oligos representing AAs 365-731 of HCV genome
CC were ligated and cloned as three separate EcoRI/BamHI subfragments
CC into the CKS fusion vector pJ0200. The sequences of these three
CC subfragments - pHCV80, pHCV77 and pHCV65 are in AAR33185-7.
CC Analysis showed that MABs H13C113 and H23C163 showed reactivity
CC pHCV80 and pHCV65.
CC (Updated on 25-MAR-2003 to correct FN field.)
XX
SQ Sequence 463 AA;
Query Match 100.0%; Score 20; DB 14; Length 463;
Best Local Similarity 100.0%; Pred. No. 3.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDFVLNNT 20
Db 407 GAPTYSWGENDTDFVLNNT 426

RESULT 12
AAB69007
ID AAB69007 standard; Peptide; 463 AA.
XX
AC AAB69007;
XX
DT 17-APR-2001 (first entry)
XX
DE HCV recombinant antigen pHCV-77 amino acid sequence SEQ ID NO:31.
XX
KW Hepatitis C virus; HCV; antigen; detection; antibody.
XX
OS Hepatitis C virus.
XX
PR US6172189-B1.
XX

```

```

XX 09-JAN-2001.
PD
XX 02-JUN-1997; 97US-0867611.
PF
XX 19-NOV-1992; 92US-0989843.
PR 10-JAN-1994; 94US-0179896.
PR 01-MAY-1996; 96US-0646757.
PR 24-AUG-1990; 90US-0572822.
PR 07-NOV-1990; 90US-0614069.
PR 21-AUG-1991; 91US-0748561.
PR 21-AUG-1991; 91US-0748566.
PR 29-OCT-1991; 91US-0748565.
XX
XX (ABBO ) ABBOTT LAB.
FA
XX Devare SG, Desai SM, Casey JM, Dailey SH, Dawson GJ, Gutierrez RA;
XX Lesniewski RR, Stewart JL, Rupprecht KR;
PI
XX WPI; 2001-122352/13.
XX
XX New recombinant antigens representing distinct antigenic regions of
PT Hepatitis C virus (HCV) genome, useful for detection of antibodies and
PT antigens in body fluids of individuals exposed to HCV -
XX
XX Example 10; Column 137-142; 167pp; English.
XX
XX The present invention describes recombinant Hepatitis C virus (HCV)
CC antigens (I). (I) is useful as a reagent for the detection of antibodies
CC and antigen in body fluids from individuals exposed to HCV. The HCV
CC assay uses reliable and efficient reagents and methods to accurately
CC detect the presence of HCV antibodies in samples obtained from
CC individuals suspected of having HCV infection. AAF32218 to AAF32235,
CC AAB51371 to AAB51379 and AAB69001 to AAB69032 represent sequences used
CC in the exemplification of the present invention.
XX
XX Sequence 463 AA;
SQ
Query Match 100.0%; Score 20; DB 22; Length 463;
Best Local Similarity 100.0%; Pred. No. 3.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDFVLNNT 20
Db 407 GAPTYSWGENDTDFVLNNT 426

RESULT 13
AAR33992
ID AAR33992 standard; Protein; 480 AA.
XX
AC AAR33992;
XX
DT 25-MAR-2003 (updated)
DT 26-JUL-1993 (first entry)
XX
DE HCV-1 E2/NS1 protein.
XX
KW Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
KW domain; immunological; cross-reactive; envelope protein; vaccine;
KW gp53 (BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.
XX
OS Synthetic.
XX
XX WO9306126-A1.
XX
PD 01-APR-1993.
XX
PF 11-SEP-1992; 92WO-US07683.
XX
PR 13-SEP-1991; '91US-0759575.
XX

```


PA (CHIR) CHIRON CORP.
 XX Houghton M, Weiner AJ;
 PI WPI; 1993-117468/14.
 XX
 XX Immuno-reactive hepatitis C virus polypeptide compens. - contg.
 PT at least 2 sequences from the first variable domain of distinct
 PT HCV isolates
 XX
 XX Disclosure; Fig 3; 106pp; English.
 XX
 CC The sequences given in AAR33992-002 represent a portion of the E2/NS1
 CC protein encoded by group I and group II HCV isolates, from amino acid
 CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
 CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
 CC 30 amino acids which shows large variation between nearly all isolates.
 CC This is an important immunoreactive domain. This putative envelope
 CC glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera
 CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
 CC flaviviruses, both of which confer protective immunity in hosts
 CC vaccinated with these polypeptides. It has been discovered that a
 CC number of important HCV epitopes vary among viral isolates and that
 CC these epitopes can be mapped to specific domains. This meant that
 CC immunologically cross-reactive polypeptides which focus on variable
 CC rather than constant domains can be produced. See also AAQ39134-48
 CC and AAR33982-91.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 XX Sequence 480 AA;
 SQ
 Query Match 100.0%; Score 20; DB 14; Length 480;
 Best Local Similarity 100.0%; Pred. No. 3.9e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 GAPTYSWGENDTDVFLNNT 20
 Db 154 GAPTYSWGENDTDVFLNNT 173
 RESULT 14
 AAE02622
 ID AAE02622 standard; Protein; 531 AA.
 XX
 XX AAE02622;
 AC
 XX
 DT 06-AUG-2001 (first entry)
 DE
 DE Chimeric HCV E2661-HBsAg S domain encoded by pCMV-II-E2661-SAg.
 XX
 XX Virus-like particle; immunogen; hepatitis B virus surface antigen;
 KW HBsAg; HBV; chimeric antigen; hepatitis C virus; HCV; virucide; vaccine;
 KW plasmid pCMV-II-E2661-SAg; E2 envelope glycoprotein; S domain.
 XX
 XX Chimeric - Hepatitis C virus.
 OS Chimeric - Hepatitis B virus.
 XX
 XX Key Location/Qualifiers
 FH Region 1..303
 FT /note= "HCV 661 E2 envelope glycoprotein"
 FT 306..531
 FT /note= "HBsAg S domain"
 XX
 XX WO200138358-A2.
 PN
 XX 31-MAY-2001.
 PD
 XX
 XX 22-NOV-2000; 2000WO-US32249.
 PF
 XX
 XX 24-NOV-1999; 99US-0167224.
 PR
 XX (CHIR) CHIRON CORP.
 PA
 XX

PI Selby M, Glazer E, Houghton M;
 XX WPI; 2001-367661/38.
 DR N-PSDB; RAD06793.
 XX
 XX Virus-like particle for use as an immunogen, comprising a first
 PT hepatitis B virus surface antigen (HBsAg) and chimeric antigen
 PT comprising a second HBsAg covalently linked to hepatitis C immunogenic
 PT polypeptide -
 XX
 XX Claim 29; Fig 4; 115pp; English.
 XX
 CC The invention relates to a virus-like particle for use as an immunogen,
 CC comprising a first hepatitis B virus surface antigen (HBsAg) and a
 CC chimeric antigen comprising a second HBsAg which is covalently linked to
 CC an hepatitis C virus (HCV) immunogenic polypeptide, where the first and
 CC the second HBsAg each comprise a substantially complete S domain.
 CC The virus-like particle is useful as immunogen and as vaccine.
 CC The present sequence is a chimeric antigen comprising HCV 661 E2
 CC envelope glycoprotein and HBsAg S domain encoded by plasmid
 CC pCMV-II-E2661-SAg.
 XX
 XX Sequence 531 AA;
 SQ
 Query Match 100.0%; Score 20; DB 22; Length 531;
 Best Local Similarity 100.0%; Pred. No. 4.3e-13;
 Matches 0; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 GAPTYSWGENDTDVFLNNT 20
 Db 165 GAPTYSWGENDTDVFLNNT 184
 RESULT 15
 AAR33185
 ID AAR33185 standard; peptide; 621 AA.
 XX
 XX AAR33185;
 AC
 XX
 DT 25-MAR-2003 (updated)
 DT 03-JUL-1993 (first entry)
 XX
 DE Sequence of subfragment pHCV80 (AAs 365-731) of the hepatitis
 DE C virus (HCV) genome.
 XX
 KW Immunogenic peptide; hepatitis C virus; immunogenic domain;
 KW monoclonal antibody; diagnosis; detection; therapy.
 XX
 XX Synthetic.
 OS
 XX WO9304205-A1.
 PN
 XX
 XX 04-MAR-1993.
 PD
 XX
 XX 21-AUG-1992; 92WO-US07189.
 PF
 XX
 XX 21-AUG-1991; 91US-0748292.
 PR
 XX (ABBO) ABBOTT LAB.
 PA
 XX Dailey SH, Desai SM, Devare SG, Johnson JE, Mehta SU;
 PI WPI; 1993-094030/11.
 XX
 XX Monoclonal antibodies specific for hepatitis C virus E2-NS1
 PT antigen - useful for diagnosis and evaluation of HCV infections
 PT and in differentiation studies
 XX
 XX Example; Pages 31-34; 48pp; English.
 PS
 XX Immunogenic domains of E2/NS1 region of HCV genome encompassing AAs
 CC 600-720 were mapped with PEPCAN analysis. Based on the EIA reactivity
 CC of a panel of HCV positive sera, peptide AAR33184 was chosen as the
 CC

CC immunogen for the generation of monoclonal antibodies to HCV NS1.
CC Several individual oligos representing AAs 365-731 of HCV genome
CC were ligated and cloned as three separate EcoRI/BamHI subfragments
CC into the CKS fusion vector pJ0200. The sequences of these three
CC subfragments - PHCV80, PHCV77 and PHCV65 are in AAR33185-7.
CC Analysis showed that MAb H13C113 and H23C163 showed reactivity
CC PHCV80 and PHCV65.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 621 AA;
Query Match 100.0%; Score 20; DB 14; Length 621;
Best Local Similarity 100.0%; Pred. No. 5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAPTYSWGENDTDVFLNNT 20
| | | | | | | | | | | | | | | | | | | | | |
Db 407 GAPTYSWGENDTDVFLNNT 426

Search completed: November 21, 2003, 20:58:05
Job time : 33.15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.6 Seconds
(without alignments)
79.832 Million cell updates/sec

Title: US-09-973-025-82

Perfect score: 20

Sequence: 1 GAPTYSWGENDTDVFLNLT 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-08-612-973-82
2	20	100.0	20	3	US-08-927-597-82
3	20	100.0	20	4	US-08-635-886C-38
4	20	100.0	50	1	US-08-262-037-43
5	20	100.0	179	3	US-08-444-818-77
6	20	100.0	402	1	US-08-460-806-13
7	20	100.0	402	1	US-08-325-630-13
8	20	100.0	403	2	US-08-483-695-39
9	20	100.0	403	2	US-07-965-285-39
10	20	100.0	403	2	US-08-487-231-39
11	20	100.0	403	3	US-09-201-912-39
12	20	100.0	463	1	US-07-748-292-9
13	20	100.0	463	3	US-08-867-611-31
14	20	100.0	463	5	PCT-US92-06965A-1
15	20	100.0	480	1	US-08-440-103-14
16	20	100.0	480	1	US-08-440-542-14
17	20	100.0	480	1	US-08-231-368-14
18	20	100.0	480	1	US-08-440-210-14
19	20	100.0	480	4	US-09-046-604-14
20	20	100.0	621	1	US-07-748-292-7
21	20	100.0	622	3	US-08-867-611-34
22	20	100.0	622	5	PCT-US92-06965A-4
23	20	100.0	663	3	US-08-824-057-3
24	20	100.0	663	4	US-09-415-582-3
25	20	100.0	663	4	US-09-693-596-4
26	20	100.0	738	3	US-08-867-611-35
27	20	100.0	738	5	PCT-US92-06965A-5

28 20 100.0 2436 3 US-08-444-818-75 Sequence 75, Appl
29 20 100.0 2772 3 US-08-444-818-89 Sequence 89, Appl
30 20 100.0 2894 2 US-08-466-975A-23 Sequence 23, Appl
31 20 100.0 2894 2 US-08-391-671A-23 Sequence 23, Appl
32 20 100.0 2894 3 US-08-467-902A-23 Sequence 23, Appl
33 20 100.0 2894 3 US-09-275-285-23 Sequence 23, Appl
34 20 100.0 2894 4 US-09-941-611-23 Sequence 23, Appl
35 20 100.0 2955 2 US-08-443-260-3 Sequence 3, Appl
36 20 100.0 2955 3 US-08-443-805A-3 Sequence 3, Appl
37 20 100.0 2955 3 US-08-443-900A-3 Sequence 3, Appl
38 20 100.0 2955 3 US-08-444-818-124 Sequence 124, App
39 20 100.0 2955 3 US-08-249-843-3 Sequence 3, Appl
40 20 100.0 2955 3 US-08-444-818-138 Sequence 138, App
41 20 100.0 3011 1 US-08-440-103-36 Sequence 36, Appl
42 20 100.0 3011 1 US-08-440-542-36 Sequence 36, Appl
43 20 100.0 3011 1 US-07-910-760-10 Sequence 10, Appl
44 20 100.0 3011 1 US-08-440-519-10 Sequence 10, Appl
45 20 100.0 3011 1 US-08-231-368-36 Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-08-612-973-82
; Sequence 82, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P. C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612.973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-612-973-82

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAPTYSWGENDTDVFLNLT 20
|||||

Db 1 GAPTYSWGENDTDVFLNNT 20

RESULT 2

US-08-927-597-82
; Sequence 82, Application US/08927597
; Patent No. 6245503

; GENERAL INFORMATION:

; APPLICANT: MAERTENS, GEERT

; APPLICANT: BOSMAN, FONS

; APPLICANT: DE MARTYNOFF, GUY

; APPLICANT: BUYSSE, MARIE-ANGE

; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

; NUMBER OF SEQUENCES: 111

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: NIXON & VANDERHYE P.C.

; STREET: 1100 NORTH GLEBE ROAD

; CITY: ARLINGTON

; STATE: VIRGINIA

; COUNTRY: U.S.A.

; ZIP: 22201-4714

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/927,597

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/612,973

; FILING DATE: 11-MAR-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: BYRNE, THOMAS E.

; REGISTRATION NUMBER: 32,205

; REFERENCE/DOCKET NUMBER: 1487-10

; TELEPHONE: (703) 816-4000

; TELEFAX: (703) 816-4100

; INFORMATION FOR SEQ ID NO: 82:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 20 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-927-597-82

Query Match 100.0%; Score 20; DB 3; Length 20;

Best Local Similarity 100.0%; Pred. No. 4.2e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDTDVFLNNT 20

Db 1 GAPTYSWGENDTDVFLNNT 20

RESULT 3

US-08-635-886C-38

; Sequence 38, Application US/08635886C

; Patent No. 6555114

; GENERAL INFORMATION:

; APPLICANT: LEROUX-ROELS, Geert

; APPLICANT: DELEYS, Robert

; APPLICANT: MAERTENS, Geert

; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C

; FILE REFERENCE: 2752-18

; CURRENT APPLICATION NUMBER: US/08/635,886C

; CURRENT FILING DATE: 1996-04-25

; PRIOR APPLICATION NUMBER: PCT/EP94/03555

; PRIOR FILING DATE: 1994-10-28

; PRIOR APPLICATION NUMBER: EP 93402718.6

; PRIOR FILING DATE: 1993-11-04

; NUMBER OF SEQ ID NOS: 286

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 38

; LENGTH: 20

; TYPE: PRT

; ORGANISM: hepatitis C virus

US-08-635-886C-38

Query Match 100.0%; Score 20; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 4.2e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDTDVFLNNT 20

Db 1 GAPTYSWGENDTDVFLNNT 20

RESULT 4

US-08-262-037-43

; Sequence 43, Application US/08262037

; Patent No. 5747239

; GENERAL INFORMATION:

; APPLICANT: Chang Yi Wang and Barbara Hosein

; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR

; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV

; NUMBER OF SEQUENCES: 136

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNEGAN

; STREET: 345 PARK AVE.

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10154

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; OPERATING SYSTEM: IBM PC COMPATIBLE

; SOFTWARE: WORDPERFECT 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/262,037

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/719,819

; FILING DATE: 24-June-1991

; APPLICATION NUMBER: 07/667,275

; FILING DATE: 11-Mar-1991

; APPLICATION NUMBER: 07/651,735

; FILING DATE: 07-Feb-1991

; APPLICATION NUMBER: 07/558,799

; FILING DATE: 26-July-1990

; APPLICATION NUMBER: 07/510,153

; FILING DATE: 16-April-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Maria C. H. Lin

; REGISTRATION NUMBER: 29,323

; REFERENCE/DOCKET NUMBER: 1151-4043 US3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-758-4800

; TELEFAX: (212) 751-6849

; TELEX: 421792

; INFORMATION FOR SEQ ID NO: 43:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 50 amino acids

; TYPE: Amino acid

; STRANDEDNESS:

; TOPOLOGY: Unknown

US-08-262-037-43

Query Match 100.0%; Score 20; DB 1; Length 50;
Best Local Similarity 100.0%; Pred. No. 8.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDFVLNNT 20
DB 4 GAPTYSWGENDTDFVLNNT 23

RESULT 5

US-08-444-818-77
; Sequence 77, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 179 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-444-818-77

Query Match 100.0%; Score 20; DB 3; Length 179;
Best Local Similarity 100.0%; Pred. No. 2.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDFVLNNT 20
DB 13 GAPTYSWGENDTDFVLNNT 32

RESULT 6

US-08-460-806-13
; Sequence 13, Application US/08460806
; Patent No. 5747241
; GENERAL INFORMATION:
; APPLICANT: MIYAMURA, TATSUO
; APPLICANT: SAITO, IZUMU
; APPLICANT: HARADA, SHIZUKO
; APPLICANT: HONDA, YOSHIKAZU
; TITLE OF INVENTION: DIAGNOSTIC REAGENT FOR HEPATITIS C
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,806
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,630
; FILING DATE: 19-OCT-1994
; APPLICATION NUMBER: US 07/956,993
; FILING DATE: 06-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5747241man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 4667-001-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR

; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-460-806-13

Query Match 100.0%; Score 20; DB 1; Length 402;
Best Local Similarity 100.0%; Pred. No. 4.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDFVLNNT 20
DB 184 GAPTYSWGENDTDFVLNNT 203

RESULT 7

US-08-325-630-13
; Sequence 13, Application US/08325630
; Patent No. 5750331
; GENERAL INFORMATION:
; APPLICANT: MIYAMURA, TATSUO
; APPLICANT: SAITO, IZUMU
; APPLICANT: HARADA, SHIZUKO
; APPLICANT: HONDA, YOSHIKAZU
; TITLE OF INVENTION: DIAGNOSTIC REAGENT FOR HEPATITIS C
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,630
; FILING DATE:

```
;
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/956,993
; FILING DATE: 06-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5750331man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 4667-001-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-325-630-13

Query Match 100.0%; Score 20; DB 1; Length 402;
Best Local Similarity 100.0%; Pred. No. 4.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDVFLNNT 20
Db 184 GAPTYSWGENDTDVFLNNT 203

RESULT 8
US-08-483-695-39
; Sequence 39, Application US/08483695
; Patent No. 5866139
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Kremesdorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSER: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,695
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/965,285
; FILING DATE: 18-MAR-1993
; APPLICATION NUMBER: FR 91 06 882
; FILING DATE: 06-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 05286-0001-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-07-965-285-39

Query Match 100.0%; Score 20; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 4.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDVFLNNT 20
Db 190 GAPTYSWGENDTDVFLNNT 209

RESULT 10
US-08-487-231-39
; Sequence 39, Application US/08487231
```

```
;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-483-695-39

Query Match 100.0%; Score 20; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 4.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDVFLNNT 20
Db 190 GAPTYSWGENDTDVFLNNT 209

RESULT 9
US-07-965-285-39
; Sequence 39, Application US/07965285
; Patent No. 5879904
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Kremesdorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSER: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/965,285
; FILING DATE: 18-MAR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 06 882
; FILING DATE: 06-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 05286-0001-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-07-965-285-39

Query Match 100.0%; Score 20; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 4.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDVFLNNT 20
Db 190 GAPTYSWGENDTDVFLNNT 209

RESULT 10
US-08-487-231-39
; Sequence 39, Application US/08487231
```

Patent No. 5919454
GENERAL INFORMATION:
APPLICANT: Brechot, Christian
APPLICANT: Kremesdorf, Dina
APPLICANT: Porchon, Colette
TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
TITLE OF INVENTION: Applications
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,231
FILING DATE: 07-JUNE-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/965,285
FILING DATE: 18-MAR-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 06 882
FILING DATE: 06-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 05286-0001-02000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 403 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-487-231-39
Query Match 100.0%; Score 20; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 4.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAPTYSWGENDTDVFLNNT 20
DB 190 GAPTYSWGENDTDVFLNNT 209
RESULT 11
US-09-201-912-39
Sequence 39, Application US/09201912
Patent No. 6210962
GENERAL INFORMATION:
APPLICANT: Brechot, Christian
APPLICANT: Kremesdorf, Dina
APPLICANT: Porchon, Colette
TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
TITLE OF INVENTION: Applications
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.

CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/201,912
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/965,285
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 05286-0001-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 403 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-201-912-39
Query Match 100.0%; Score 20; DB 3; Length 403;
Best Local Similarity 100.0%; Pred. No. 4.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAPTYSWGENDTDVFLNNT 20
DB 190 GAPTYSWGENDTDVFLNNT 209
RESULT 12
US-07-748-292-9
Sequence 9, Application US/07748292
Patent No. 5308750
GENERAL INFORMATION:
APPLICANT: MEHTA, SMRITI U.
APPLICANT: JOHNSON, JILL E.
APPLICANT: DAILEY, STEPHEN H.
APPLICANT: DESAI, SURESH M.
APPLICANT: DEVARE, SUSHIL G.
TITLE OF INVENTION: MONOCLONAL ANTIBODIES TO PUTATIVE HCV
TITLE OF INVENTION: E2/NS1 PROTEINS AND METHODS FOR USING SAME
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES D-377/AP6D
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: ILLINOIS
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/748,292
FILING DATE: 19910821
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/456,162
FILING DATE: 22-DEC-1989
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/610,180
FILING DATE: 07-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: FOREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 4767.US.P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-748-292-9

Query Match 100.0%; Score 20; DB 1; Length 463;
Best Local Similarity 100.0%; Pred. No. 5.2e-13; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAPTYSWGENDTDFVLNNT 20
DB 407 GAPTYSWGENDTDFVLNNT 426

RESULT 13
US-08-867-611-31
Sequence 31, Application US/08867611
Patent No. 6172189
GENERAL INFORMATION:
APPLICANT: DEVARE, SUSHIL G
APPLICANT: DESAI, SURESH M
APPLICANT: CASEY, JAMES M
APPLICANT: DAILEY, STEPHEN H
APPLICANT: DAWSON, GEORGE J
APPLICANT: GUTIERREZ, ROBIN A
APPLICANT: LESNIEWSKI, RICHARD R
APPLICANT: STEWART, JAMES L
APPLICANT: RUPPRECHT, KEVIN R
TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT
TITLE OF INVENTION: ANTIGENS
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES
STREET: ONE ABBOTT PARK ROAD, CHAD377/AP6D2
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: 02-JUN-1997
APPLICATION NUMBER: US/08/867,611
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
FILING DATE: US/08/646,757
APPLICATION NUMBER: US/08/179,896
FILING DATE:
PRIOR APPLICATION DATA:
FILING DATE: US 07/572,822
APPLICATION NUMBER: US 07/614,069
FILING DATE: 07-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/748,561

FILING DATE: 21-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/748,565
FILING DATE: 21-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/748,566
FILING DATE: 21-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: FOREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 4834.US.P6
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-867-611-31

Query Match 100.0%; Score 20; DB 3; Length 463;
Best Local Similarity 100.0%; Pred. No. 5.2e-13; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAPTYSWGENDTDFVLNNT 20
DB 407 GAPTYSWGENDTDFVLNNT 426

RESULT 14
PCT-US92-06965A-1
Sequence 1, Application PC/TUS9206965A
GENERAL INFORMATION:
APPLICANT: DEVARE, S.
APPLICANT: DESAI, S.
APPLICANT: DAILEY, S.
TITLE OF INVENTION: HCV SYNTHETIC PEPTIDE FROM NS1 REGION
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: ILLINOIS
COUNTRY: U.S.
ZIP: 60065-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/06965A
FILING DATE: 19920821
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: FOREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 4834PC.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US92-06965A-1

Query Match 100.0%; Score 20; DB 5; Length 463;
Best Local Similarity 100.0%; Pred. No. 5.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDVFLVNT 20
DB 407 GAPTYSWGENDTDVFLVNT 426

RESULT 15

US-08-440-103-14
; Sequence 14, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,103
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-103-14

Query Match 100.0%; Score 20; DB 1; Length 480;
Best Local Similarity 100.0%; Pred. No. 5.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDVFLVNT 20
DB 154 GAPTYSWGENDTDVFLVNT 173

Search completed: November 21, 2003, 21:15:16
Job time : 10.6 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time 19.5 Seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-82

Perfect score: 20

Sequence: 1 GAPTYSWGENDTDFVLNLT 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	10	US-09-973-025-82
2	20	100.0	20	11	US-09-899-303-82
3	20	100.0	20	11	US-09-995-808-82
4	20	100.0	20	11	US-09-995-860-82
5	20	100.0	20	12	US-09-995-791-82
6	20	100.0	254	10	US-09-407-430-3
7	20	100.0	363	10	US-09-407-430-2
8	20	100.0	637	12	US-10-187-257-4
9	20	100.0	637	12	US-10-265-083-2
10	20	100.0	2894	10	US-09-941-611-23
11	20	100.0	2894	15	US-10-044-985-23
12	20	100.0	3011	9	US-09-916-359-2
13	20	100.0	3011	16	US-10-232-643-6
14	12	60.0	2985	15	US-10-259-275-40
15	11	55.0	36	12	US-10-351-641-1084

16	11	55.0	37	10	US-09-921-397-4	Sequence 4, Appli
17	11	55.0	96	10	US-09-921-397-82	Sequence 82, Appl
18	11	55.0	176	10	US-09-921-397-81	Sequence 81, Appl
19	11	55.0	250	10	US-09-952-572-8	Sequence 8, Appli
20	11	55.0	350	10	US-09-929-955-4	Sequence 4, Appli
21	11	55.0	350	14	US-10-104-966-4	Sequence 4, Appli
22	11	55.0	363	12	US-10-128-587A-97	Sequence 97, Appl
23	11	55.0	363	15	US-10-128-590-97	Sequence 97, Appl
24	11	55.0	3011	9	US-09-742-659-4	Sequence 4, Appli
25	11	55.0	3011	10	US-09-238-076-20	Sequence 20, Appl
26	11	55.0	3011	10	US-09-952-572-9	Sequence 9, Appli
27	11	55.0	3011	10	US-09-929-955-1	Sequence 1, Appli
28	11	55.0	3011	10	US-09-747-419-20	Sequence 20, Appl
29	11	55.0	3011	11	US-09-891-894-3	Sequence 3, Appli
30	11	55.0	3011	11	US-09-995-937-20	Sequence 20, Appl
31	11	55.0	3011	11	US-09-917-563-20	Sequence 20, Appl
32	11	55.0	3011	12	US-10-184-150-3	Sequence 3, Appli
33	11	55.0	3011	14	US-10-104-966-1	Sequence 1, Appli
34	11	55.0	3011	15	US-10-259-275-20	Sequence 20, Appl
35	11	55.0	3012	10	US-09-238-076-2	Sequence 2, Appli
36	11	55.0	3012	11	US-09-995-937-2	Sequence 2, Appli
37	11	55.0	3012	11	US-09-917-563-2	Sequence 2, Appli
38	10	50.0	31	10	US-09-921-397-15	Sequence 15, Appl
39	8	40.0	20	10	US-09-973-025-81	Sequence 81, Appl
40	8	40.0	20	11	US-09-899-303-81	Sequence 81, Appl
41	8	40.0	20	11	US-09-995-808-81	Sequence 81, Appl
42	8	40.0	20	11	US-09-995-860-81	Sequence 81, Appl
43	8	40.0	20	12	US-09-995-791-81	Sequence 81, Appl
44	8	40.0	28	10	US-09-921-397-6	Sequence 6, Appli
45	8	40.0	43	10	US-09-921-397-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1

US-09-973-025-82
; Sequence 82, Application US/09973025
; Publication No. US20020182706A1

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

BOSMAN, FONS

DE MARTYNOFF, GUY

BUYSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHUYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/973,025

FILING DATE: 10-Oct-2001

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 82:
US-09-973-025-82

Query Match 100.0%; Score 20; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.1e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDVFLNNT 20
|||||
DB 1 GAPTYSWGENDTDVFLNNT 20

RESULT 2
US-09-899-303-82
; Sequence 82, Application US/09899303
; Publication No. US20030036110A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA: US/09/899,303
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 82:
US-09-899-303-82

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.1e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDVFLNNT 20
|||||

Db 1 GAPTYSWGENDTDVFLNNT 20

RESULT 3
US-09-995-808-82
; Sequence 82, Application US/09995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 82
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-82

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.1e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDVFLNNT 20
|||||
DB 1 GAPTYSWGENDTDVFLNNT 20

RESULT 4
US-09-995-860-82
; Sequence 82, Application US/09995860
; Publication No. US20030118603A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-89
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 82
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-82

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.1e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDVFLNNT 20
|||||
DB 1 GAPTYSWGENDTDVFLNNT 20

RESULT 5
US-09-995-791-82
; Sequence 82, Application US/09995791
; Publication No. US20030147918A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 82

```
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-82

Query Match      100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.1e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDDTDVFLNNT 20
   |||||
Db 1 GAPTYSWGENDDTDVFLNNT 20
   |||||

RESULT 6
US-09-407-430-3
; Sequence 3, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-407-430-3

Query Match      100.0%; Score 20; DB 10; Length 254;
Best Local Similarity 100.0%; Pred. No. 7.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDDTDVFLNNT 20
   |||||
Db 118 GAPTYSWGENDDTDVFLNNT 137
   |||||

RESULT 7
US-09-407-430-2
; Sequence 2, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-09-407-430-2

Query Match      100.0%; Score 20; DB 10; Length 363;
Best Local Similarity 100.0%; Pred. No. 1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDDTDVFLNNT 20
   |||||
Db 140 GAPTYSWGENDDTDVFLNNT 159
   |||||

RESULT 8
US-10-187-257-4
; Sequence 4, Application US/10187257
; Publication No. US20030138458A1
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: O'HAGAN, Derek
; TITLE OF INVENTION: HCV ELE2 VACCINE COMPOSITIONS
; FILE REFERENCE: 2302-17206
; CURRENT APPLICATION NUMBER: US/10/187,257
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
; OTHER INFORMATION: amino acid
US-10-187-257-4

Query Match      100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDDTDVFLNNT 20
   |||||
Db 351 GAPTYSWGENDDTDVFLNNT 370
   |||||

RESULT 9
US-10-265-083-2
; Sequence 2, Application US/10265083
; Publication No. US20030170273A1
; GENERAL INFORMATION:
; APPLICANT: O'HAGAN, Derek
; APPLICANT: VALIANTE, Nicholas
; TITLE OF INVENTION: ADJUVANT COMPOSITIONS
; FILE REFERENCE: 2302-18357.30
; CURRENT APPLICATION NUMBER: US/10/265,083
; CURRENT FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7
US-10-265-083-2

Query Match      100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDDTDVFLNNT 20
   |||||
Db 351 GAPTYSWGENDDTDVFLNNT 370
   |||||

RESULT 10
US-09-941-611-23
; Sequence 23, Application US/09941611
; Patent No. US20020106640A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; ANTIBODIES TO HEPATITIS C VIRUS
```

NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/941,611
FILING DATE: 30-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: 1995-02-21
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-941-611-23

Query Match 100.0%; Score 20; DB 10; Length 2894;
Best Local Similarity 100.0%; Pred. No. 6.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDVFLNNT 20
|||||
DB 523 GAPTYSWGENDTDVFLNNT 542

RESULT 11
US-10-044-995-23
Sequence 23, Application US/10044995
Publication No. US20030049685A1
GENERAL INFORMATION:
APPLICANT: DELEYS, ROBERT J
POLLET, DIRK
MAERTENS, GEERT
VAN HEUVERSWUN, HUGO
TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
ANTIBODIES TO HEPATITIS C VIRUS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/044,995
FILING DATE: 15-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: <Unknown>
APPLICATION NUMBER: US 07/920,286
FILING DATE: 14-OCT-1992
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-044-995-23

Query Match 100.0%; Score 20; DB 15; Length 2894;
Best Local Similarity 100.0%; Pred. No. 6.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDVFLNNT 20
|||||
DB 523 GAPTYSWGENDTDVFLNNT 542

RESULT 12
US-09-916-359-2
Sequence 2, Application US/09916359
Patent No. US20020034734A1
GENERAL INFORMATION:
APPLICANT: Veronique Barban
TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
TREATING C HEPATITIS
FILE REFERENCE: PMCF97-03A
CURRENT APPLICATION NUMBER: US/09/916,359
CURRENT FILING DATE: 2001-07-26
PRIOR APPLICATION NUMBER: 09/388,874
PRIOR FILING DATE: 1999-09-02
PRIOR APPLICATION NUMBER: 97/02,887
PRIOR FILING DATE: 1997-03-06
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 3011
TYPE: PRT
ORGANISM: Virus
US-09-916-359-2

Query Match 100.0%; Score 20; DB 9; Length 3011;
Best Local Similarity 100.0%; Pred. No. 6.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDVFLNNT 20
|||||

Db 523 GAPTYSWGENDDTDVFLNNT 542

RESULT 13

US-10-232-643-6

Sequence 6, Application US/10232643

Publication No. US20030129586A1

GENERAL INFORMATION:

APPLICANT: HOUGHTON, MICHAEL

CHOO, QUI-LIM

HAN, JANG

CHOE, JOONHO

TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING

HELICASE ACTIVITY AND IMPROVED SOLUBILITY

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: CHIRON CORPORATION

STREET: Intellectual Property - R440, P.O. Box 8097

CITY: Emeryville

STATE: California

COUNTRY: U.S.A.

ZIP: 94662-8097

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/10/232,643

FILING DATE: 30-Aug-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/483,799

FILING DATE: <Unknown>

APPLICATION NUMBER: 08/529,169

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Harbin, Alisa A.

REGISTRATION NUMBER: 33,895

REFERENCE/DOCKET NUMBER: 0100.005

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 923-3274

TELEFAX: (510) 655-3542

TELEX: n/a

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 3011 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Duplication

LOCATION: 9

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Lys or Arg"

FEATURE:

NAME/KEY: Duplication

LOCATION: 11

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Asn or Thr"

FEATURE:

NAME/KEY: Duplication

LOCATION: 176

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Ile or Thr"

FEATURE:

NAME/KEY: Duplication

LOCATION: 334

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Met or Val"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2502

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Leu or Phe"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1454

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Cys or Tyr"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1471

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Thr or Ser"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1877

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Glu or Gly"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1948

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Leu or His"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1949

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Ser or Cys"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2021

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Gly or Val"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2349

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Thr or Ser"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2385

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Tyr or Phe"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2386

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Ser or Ala"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2502

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Leu or Phe"

FEATURE:

NAME/KEY: Duplication

LOCATION: 603

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Leu or Ile"

FEATURE:

NAME/KEY: Duplication

LOCATION: 848

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Tyr or Asn"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1114

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Pro or Ser"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1117

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Ser or Thr"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1276

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Pro or Leu"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1454

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Cys or Tyr"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1471

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Thr or Ser"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1877

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Glu or Gly"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1948

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Leu or His"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1949

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Ser or Cys"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2021

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Gly or Val"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2349

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Thr or Ser"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2385

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Tyr or Phe"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2386

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Ser or Ala"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2502

OTHER INFORMATION: /note= "There exists a

heterogeneity at this position - Xaa = Leu or Phe"

```
;
;
; NAME/KEY: Duplication
; LOCATION: 2690
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Arg or Gly"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2921
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Arg or Gly"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2996
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Leu or Pro"
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-232-643-6

Query Match      100.0%; Score 20; DB 16; Length 3011;
Best Local Similarity 100.0%; Pred. No. 6.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDTDVFLNNT 20
Db 523 GAPTYSWGENDTDVFLNNT 542
|||||

RESULT 14
US-10-259-275-40
; Sequence 40, Application US/10259275
; Publication No. US20030125541A1
; GENERAL INFORMATION:
; APPLICANT: Lemon, Stanley M.
; APPLICANT: Yi, Minkyung
; TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE
; FILE REFERENCE: 265.0007 0120
; CURRENT APPLICATION NUMBER: US/10/259,275
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: US 60/171,909
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: US 09/747,419
; PRIOR FILING DATE: 2000-12-23
; PRIOR APPLICATION NUMBER: US 60/325,236
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/338,123
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 40
; LENGTH: 2985
; TYPE: PRT
; ORGANISM: ARTIFICIAL
; FEATURE:
; OTHER INFORMATION: amino acid sequence encoded by nucleotides 2077-11121 of SEQ ID
; OTHER INFORMATION: NO:39
US-10-259-275-40

Query Match      60.0%; Score 12; DB 15; Length 2985;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 PTYSWGENDTDV 14
Db 526 PTYSWGENDTDV 537
|||||

RESULT 15
US-10-351-641-1084
; Sequence 1084, Application US/10351641
; Publication No. US20030186874A1
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
```

```
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
; APPLICANT: Lambert, D.
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; TITLE OF INVENTION: PROPERTIES
; FILE REFERENCE: 7872-100
; CURRENT APPLICATION NUMBER: US/10/351,641
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: 09/350,641
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/315,304
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1757
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1084
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-10-351-641-1084

Query Match      55.0%; Score 11; DB 12; Length 36;
Best Local Similarity 100.0%; Pred. No. 5.4e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 NDTDFVLNNT 20
Db 17 NDTDFVLNNT 27
|||||

Search completed: November 21, 2003, 22:19:39
Job time : 19.5 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 163.1 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-82

Perfect score: 20
Sequence: 1 GAPTYSWGENDTDFVLNNT 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0
Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA_Main:*

1: /cgn2_6/ptodata/1/paa/PTUS COMB.pcp.*
2: /cgn2_6/ptodata/1/paa/US06 COMB.pcp.*
3: /cgn2_6/ptodata/1/paa/US07 COMB.pcp.*
4: /cgn2_6/ptodata/1/paa/US08 COMB.pcp.*
5: /cgn2_6/ptodata/1/paa/US081 COMB.pcp.*
6: /cgn2_6/ptodata/1/paa/US082 COMB.pcp.*
7: /cgn2_6/ptodata/1/paa/US083 COMB.pcp.*
8: /cgn2_6/ptodata/1/paa/US084 COMB.pcp.*
9: /cgn2_6/ptodata/1/paa/US085 COMB.pcp.*
10: /cgn2_6/ptodata/1/paa/US086 COMB.pcp.*
11: /cgn2_6/ptodata/1/paa/US087 COMB.pcp.*
12: /cgn2_6/ptodata/1/paa/US088 COMB.pcp.*
13: /cgn2_6/ptodata/1/paa/US089 COMB.pcp.*
14: /cgn2_6/ptodata/1/paa/US090 COMB.pcp.*
15: /cgn2_6/ptodata/1/paa/US091 COMB.pcp.*
16: /cgn2_6/ptodata/1/paa/US092 COMB.pcp.*
17: /cgn2_6/ptodata/1/paa/US093 COMB.pcp.*
18: /cgn2_6/ptodata/1/paa/US094 COMB.pcp.*
19: /cgn2_6/ptodata/1/paa/US095 COMB.pcp.*
20: /cgn2_6/ptodata/1/paa/US096 COMB.pcp.*
21: /cgn2_6/ptodata/1/paa/US097A COMB.pcp.*
22: /cgn2_6/ptodata/1/paa/US097B COMB.pcp.*
23: /cgn2_6/ptodata/1/paa/US098 COMB.pcp.*
24: /cgn2_6/ptodata/1/paa/US099A COMB.pcp.*
25: /cgn2_6/ptodata/1/paa/US099B COMB.pcp.*
26: /cgn2_6/ptodata/1/paa/US100 COMB.pcp.*
27: /cgn2_6/ptodata/1/paa/US101 COMB.pcp.*
28: /cgn2_6/ptodata/1/paa/US102 COMB.pcp.*
29: /cgn2_6/ptodata/1/paa/US103 COMB.pcp.*
30: /cgn2_6/ptodata/1/paa/US104 COMB.pcp.*
31: /cgn2_6/ptodata/1/paa/US106 COMB.pcp.*
32: /cgn2_6/ptodata/1/paa/US60 COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	13	US-08-974-690C-38
					Sequence 38, Appl

2	20	100.0	20	23	US-09-899-303-82	Sequence 82, Appl
3	20	100.0	20	23	US-09-899-303A-82	Sequence 82, Appl
4	20	100.0	20	25	US-09-973-025-82	Sequence 82, Appl
5	20	100.0	20	25	US-09-995-791-82	Sequence 82, Appl
6	20	100.0	20	25	US-09-995-808-82	Sequence 82, Appl
7	20	100.0	20	25	US-09-995-860-82	Sequence 82, Appl
8	20	100.0	20	26	US-10-020-510-82	Sequence 82, Appl
9	20	100.0	20	29	US-10-321-798-82	Sequence 82, Appl
10	20	100.0	50	8	US-08-475-482-43	Sequence 43, Appl
11	20	100.0	50	8	US-08-477-072-43	Sequence 43, Appl
12	20	100.0	50	8	US-08-477-582-43	Sequence 43, Appl
13	20	100.0	50	8	US-08-480-253-43	Sequence 43, Appl
14	20	100.0	122	8	US-08-436-966-5	Sequence 5, Appl
15	20	100.0	179	8	US-08-403-590B-77	Sequence 77, Appl
16	20	100.0	179	8	US-08-444-112-77	Sequence 77, Appl
17	20	100.0	254	1	PCT-US00-26395-3	Sequence 3, Appl
18	20	100.0	254	18	US-09-407-430-3	Sequence 3, Appl
19	20	100.0	333	30	US-10-445-724-3	Sequence 3, Appl
20	20	100.0	333	32	US-60-409-909-4	Sequence 4, Appl
21	20	100.0	363	1	PCT-US00-26395-2	Sequence 2, Appl
22	20	100.0	363	18	US-09-407-430-2	Sequence 2, Appl
23	20	100.0	416	8	US-08-436-966-9	Sequence 9, Appl
24	20	100.0	463	3	US-07-748-561-1	Sequence 1, Appl
25	20	100.0	463	3	US-07-989-843-31	Sequence 31, Appl
26	20	100.0	463	8	US-08-463-849-31	Sequence 31, Appl
27	20	100.0	463	8	US-08-463-884-31	Sequence 31, Appl
28	20	100.0	463	20	US-09-690-359-31	Sequence 31, Appl
29	20	100.0	480	8	US-08-471-498-14	Sequence 14, Appl
30	20	100.0	531	21	US-09-721-480-7	Sequence 7, Appl
31	20	100.0	622	3	US-07-748-561-4	Sequence 4, Appl
32	20	100.0	622	3	US-07-989-843-34	Sequence 34, Appl
33	20	100.0	622	8	US-08-463-849-34	Sequence 34, Appl
34	20	100.0	622	8	US-08-463-884-34	Sequence 34, Appl
35	20	100.0	622	20	US-09-690-359-34	Sequence 34, Appl
36	20	100.0	637	1	PCT-US02-20676-4	Sequence 4, Appl
37	20	100.0	637	1	PCT-US02-30423-2	Sequence 2, Appl
38	20	100.0	637	1	PCT-US02-31486-2	Sequence 2, Appl
39	20	100.0	637	27	US-10-187-257-4	Sequence 4, Appl
40	20	100.0	637	28	US-10-265-083-2	Sequence 2, Appl
41	20	100.0	663	6	US-08-282-959-3	Sequence 3, Appl
42	20	100.0	663	14	US-09-073-406-4	Sequence 4, Appl
43	20	100.0	663	29	US-10-371-040-4	Sequence 4, Appl
44	20	100.0	732	3	US-07-748-561-5	Sequence 5, Appl
45	20	100.0	738	3	US-07-989-843-35	Sequence 35, Appl

ALIGNMENTS

RESULT 1
US-08-974-690C-38
; Sequence 38, Application US/08974690C
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/08/974,690C
; CURRENT FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-974-690C-38

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;

Qy 1 GAPTYSWGENDTDVFLNNT 20
Db 1 GAPTYSWGENDTDVFLNNT 20

RESULT 2

US-09-899-303-82
; Sequence 82, Application US/09899303
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 82:

US-09-899-303-82
Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;

Qy 1 GAPTYSWGENDTDVFLNNT 20
Db 1 GAPTYSWGENDTDVFLNNT 20

RESULT 3

US-09-899-303A-82
; Sequence 82, Application US/09899303A
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025

Qy 1 GAPTYSWGENDTDVFLNNT 20
Db 1 GAPTYSWGENDTDVFLNNT 20

RESULT 4

US-09-973-025-82
; Sequence 82, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303A
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 82:

US-09-899-303A-82
Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;

Qy 1 GAPTYSWGENDTDVFLNNT 20
Db 1 GAPTYSWGENDTDVFLNNT 20

US-09-973-025-82
; Sequence 82, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025

Qy 1 GAPTYSWGENDTDVFLNNT 20
Db 1 GAPTYSWGENDTDVFLNNT 20

US-09-973-025-82
; Sequence 82, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025

Qy 1 GAPTYSWGENDTDVFLNNT 20
Db 1 GAPTYSWGENDTDVFLNNT 20

US-09-973-025-82
; Sequence 82, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025

Qy 1 GAPTYSWGENDTDVFLNNT 20
Db 1 GAPTYSWGENDTDVFLNNT 20

US-09-973-025-82
; Sequence 82, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025

; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 82:
US-09-973-025-82

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDTDFVLNNT 20
| | | | | | | | | | | | | | | | | |
Db 1 GAPTYSWGENDTDFVLNNT 20

RESULT 5
US-09-995-791-82
; Sequence 82, Application US/09995791
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 82
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-82

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDTDFVLNNT 20
| | | | | | | | | | | | | | | | | |
Db 1 GAPTYSWGENDTDFVLNNT 20

RESULT 6
US-09-995-808-82
; Sequence 82, Application US/09995808
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 82
; LENGTH: 20
; TYPE: PRT

; ORGANISM: Hepatitis C virus
US-09-995-808-82

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDTDFVLNNT 20
| | | | | | | | | | | | | | | | | |
Db 1 GAPTYSWGENDTDFVLNNT 20

RESULT 7
US-09-995-860-82
; Sequence 82, Application US/09995860
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 82
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-82

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDTDFVLNNT 20
| | | | | | | | | | | | | | | | | |
Db 1 GAPTYSWGENDTDFVLNNT 20

RESULT 8
US-10-020-510-82
; Sequence 82, Application US/10020510
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-72
; CURRENT APPLICATION NUMBER: US/10/020,510
; CURRENT FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 82
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-020-510-82

Query Match 100.0%; Score 20; DB 26; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDTDFVLNNT 20
| | | | | | | | | | | | | | | | | |
Db 1 GAPTYSWGENDTDFVLNNT 20

RESULT 9
US-10-321-798-82
; Sequence 82, Application US/10321798
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-79
; CURRENT APPLICATION NUMBER: US/10/321,798
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 82
; LENGTH: 20
; TYPE: PRT

```
; FILE REFERENCE: 2551-93
; CURRENT APPLICATION NUMBER: US/10/321,798
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: 60/418,358
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: 10/020,510
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: Patentin 3.1
; SEQ ID NO 82
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-321-798-82

Query Match      100.0%; Score 20; DB 29; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDVFLNNT 20
Db 1 GAPTYSWGENDTDVFLNNT 20

RESULT 10
US-08-475-482-43
; Sequence 43, Application US/08475482
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,482
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 17-June-1994
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-10-321-798-82

Query Match      100.0%; Score 20; DB 29; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDVFLNNT 20
Db 1 GAPTYSWGENDTDVFLNNT 20

RESULT 11
US-08-477-072-43
; Sequence 43, Application US/08477072
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,072
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 17-June-1994
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-477-072-43

Query Match      100.0%; Score 20; DB 8; Length 50;
Best Local Similarity 100.0%; Pred. No. 4.4e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDVFLNNT 20
```

```
Db      4 GAPTYSWGENDTDVFLNNT 23

RESULT 12
US-08-477-582-43
; Sequence 43, Application US/08477582
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,582
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 07-June-1995
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US7
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-477-582-43

Query Match      100.0%; Score 20; DB 8; Length 50;
Best Local Similarity 100.0%; Pred. No. 4.4e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAPTYSWGENDTDVFLNNT 20
Db      4 GAPTYSWGENDTDVFLNNT 23

RESULT 13
US-08-480-253-43
; Sequence 43, Application US/08480253
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
```

```
;
;
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,253
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 07-June-1995
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US9
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-480-253-43

Query Match      100.0%; Score 20; DB 8; Length 50;
Best Local Similarity 100.0%; Pred. No. 4.4e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 GAPTYSWGENDTDVFLNNT 20
Db      4 GAPTYSWGENDTDVFLNNT 23

RESULT 14
US-08-436-966-5
; Sequence 5, Application US/08436966
; GENERAL INFORMATION:
; APPLICANT: Miyamura, Tatsuo
; APPLICANT: Saito, Izumi
; APPLICANT: Houghton, Michael
; APPLICANT: Weiner, Amy J.
; APPLICANT: Han, Jang
; APPLICANT: Kolberg, Janice A.
; APPLICANT: Cha, Tai-An
; APPLICANT: Irvine, Bruce D.
; TITLE OF INVENTION: New HCV Isolates
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
```

STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/436.966
FILING DATE: 08-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/456.142
FILING DATE: 21-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/408.045
ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0089.007
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-3274
TELEFAX: 510-855-3542
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 122 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-436-966-5

Query Match 100.0%; Score 20; DB 8; Length 122;
Best Local Similarity 100.0%; Pred. No. 9.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDTDVFLNNT 20
|||||
Db 92 GAPTYSWGENDTDVFLNNT 111

RESULT 15
US-08-403-590B-77
Sequence 77, Application US/08403590B
GENERAL INFORMATION:
APPLICANT: Chien, David Y.
APPLICANT: Rutter, William J.
TITLE OF INVENTION: NANEV Diagnostics and Vaccines
NUMBER OF SEQUENCES: 777
CORRESPONDENCE ADDRESS:
ADDRESSER: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403.590B
FILING DATE: 14-MAR-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0110.002
TELECOMMUNICATION INFORMATION:

TELEPHONE: (508)359-3876
TELEFAX: (508)359-3885
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 179 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-403-590B-77

Query Match 100.0%; Score 20; DB 8; Length 179;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDTDVFLNNT 20
|||||
Db 13 GAPTYSWGENDTDVFLNNT 32

Search completed: November 21, 2003, 22:09:52
Job time : 163.1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.55 Seconds
(without alignments)
117.536 Million cell updates/sec

Title: US-09-973-025-82

Perfect score: 20

Sequence: 1 GAPTYSWGENDTVFVLNLT 20

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- 1: /cgm2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
- 2: /cgm2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
- 3: /cgm2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 4: /cgm2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 5: /cgm2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 6: /cgm2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
- 7: /cgm2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	6	US-10-651-165-38
2	20	100.0	333	1	PCT-US03-19834-3
3	20	100.0	333	6	US-10-655-562-4
4	20	100.0	637	1	PCT-US03-33610-4
5	20	100.0	3011	1	PCT-US03-19834-2
6	11	55.0	30	6	US-10-296-734-478
7	11	55.0	36	5	US-09-350-841A-1084
8	11	55.0	36	5	US-09-350-369D-1703
9	11	55.0	36	5	US-09-350-369E-1703
10	11	55.0	36	6	US-10-351-641-1084
11	11	55.0	2010	6	US-10-296-734-814
12	11	55.0	3011	6	US-10-296-734-406
13	11	55.0	5985	6	US-10-296-734-810
14	10	50.0	10	6	US-10-440-390-38
15	8	40.0	20	6	US-10-651-165-37
16	8	40.0	30	6	US-10-296-734-476
17	8	40.0	2011	6	US-10-296-734-812
18	6	30.0	157	6	US-10-679-063-20525
19	6	30.0	177	6	US-10-425-114A-50431
20	6	30.0	384	6	US-10-425-114A-36814
21	6	30.0	394	6	US-10-425-114A-51661
22	6	30.0	414	6	US-10-425-114A-56123
23	6	30.0	450	6	US-10-425-114A-68760
24	6	30.0	450	6	US-10-425-114A-41442
25	6	30.0	597	1	PCT-US03-26780-1791
26	6	30.0	609	6	US-10-679-063-9124

27	6	30.0	609.	6	US-10-679-063-13280	Sequence 13280, A
28	6	30.0	648	6	US-10-425-114A-62408	Sequence 62408, A
29	6	30.0	2329	5	US-09-897-516A-4699	Sequence 4699, A
30	5	25.0	9	1	PCT-US02-27760-114	Sequence 114, App
31	5	25.0	9	1	PCT-US02-27760-329	Sequence 329, App
32	5	25.0	9	1	PCT-US02-27760-332	Sequence 332, App
33	5	25.0	10	1	PCT-US02-27760-161	Sequence 161, App
34	5	25.0	10	1	PCT-US02-27760-259	Sequence 259, App
35	5	25.0	10	1	PCT-US02-27760-355	Sequence 355, App
36	5	25.0	10	1	PCT-US02-27760-374	Sequence 374, App
37	5	25.0	10	4	US-08-461-651A-173	Sequence 173, App
38	5	25.0	10	4	US-08-461-651A-174	Sequence 174, App
39	5	25.0	10	4	US-08-461-651A-175	Sequence 175, App
40	5	25.0	10	4	US-08-461-651A-176	Sequence 176, App
41	5	25.0	10	4	US-08-461-651A-177	Sequence 177, App
42	5	25.0	10	4	US-08-461-651A-178	Sequence 178, App
43	05	254.0	62	5	US-09-897-516A-5073	Sequence 5073, Ap
44	5	25.0	66	6	US-10-425-114A-54201	Sequence 54201, A
45	5	25.0	77	6	US-10-425-114A-50082	Sequence 50082, A

ALIGNMENTS

RESULT 1

US-10-651-165-38
; Sequence 38, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus

US-10-651-165-38

Query Match 100.0%; Score 20; DB 6; Length 20;

Best Local Similarity 100.0%; Pred. No. 4.7e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GAPTYSWGENDTVFVLNLT 20

Db 1 GAPTYSWGENDTVFVLNLT 20

RESULT 2

PCT-US03-19834-3

; Sequence 3, Application PC/TUS0319834

; GENERAL INFORMATION:

; APPLICANT: STAPLETON, JACK T.

; APPLICANT: WUENSCHMANN, SABINA

; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND

; TITLE OF INVENTION: METHODS FOR TREATING HYPERLIPIDEMIA

; FILE REFERENCE: IOWA:045WO

; CURRENT APPLICATION NUMBER: PCT/US03/19834

; CURRENT FILING DATE: 2003-05-24

; PRIOR APPLICATION NUMBER: 10/445,724

; PRIOR FILING DATE: 2003-05-27

; PRIOR APPLICATION NUMBER: 60/392,158

```
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-3

Query Match      100.0%; Score 20; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAPTYSWGENDTDFVLNNT 20
Db      141 GAPTYSWGENDTDFVLNNT 160

RESULT 3
US-10-655-562-4
; Sequence 4, Application US/10655562
; GENERAL INFORMATION:
; APPLICANT: BISHOP-HURLEY, SHARON L.
; APPLICANT: SCHMIDT, FRANCIS J.
; APPLICANT: SMITH, ARNOLD L.
; TITLE OF INVENTION: PHAGE-DISPLAY PEPTIDES AS NOVEL ANTIMICROBIAL AGENTS
; FILE REFERENCE: UOVO:022US
; CURRENT APPLICATION NUMBER: US/10/655,562
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: 60,409,909
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-10-655-562-4

Query Match      100.0%; Score 20; DB 6; Length 333;
Best Local Similarity 100.0%; Pred. No. 3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAPTYSWGENDTDFVLNNT 20
Db      141 GAPTYSWGENDTDFVLNNT 160

RESULT 4
PCT-US03-33610-4
; Sequence 4, Application PC/TUS0333610
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: SELBY, Mark
; APPLICANT: PALIARD, Xavier
; TITLE OF INVENTION: ACTIVATION OF HCV-SPECIFIC T-CELLS
; FILE REFERENCE: 2300-1612.60
; CURRENT APPLICATION NUMBER: PCT/US03/33610
; CURRENT FILING DATE: 2003-10-24
; PRIOR APPLICATION NUMBER: 10/281,341
; PRIOR FILING DATE: 2002-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
PCT-US03-33610-4

Query Match      100.0%; Score 20; DB 1; Length 637;
Best Local Similarity 100.0%; Pred. No. 4.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAPTYSWGENDTDFVLNNT 20
Db      351 GAPTYSWGENDTDFVLNNT 370

RESULT 5
PCT-US03-19834-2
; Sequence 2, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; FILE REFERENCE: IOWA:04SWO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-2

Query Match      100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.3e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAPTYSWGENDTDFVLNNT 20
Db      523 GAPTYSWGENDTDFVLNNT 542

RESULT 6
US-10-296-734-478
; Sequence 478, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 478
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 36
US-10-296-734-478

Query Match      55.0%; Score 11; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 6.8e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      10 NDTDFVLNNT 20
Db      9 NDTDFVLNNT 19

; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
PCT-US03-33610-4
```



```
RESULT 7
US-09-350-841A-1084
; Sequence 1084, Application US/09350841A
; GENERAL INFORMATION:
; APPLICANT: Jeffs, Peter;
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION OF MEMBRANE
; FILE REFERENCE: 7872-066-999
; CURRENT APPLICATION NUMBER: US/09/350,841A
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 1946
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 1084
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-09-350-841A-1084

Query Match      55.0%; Score 11; DB 5; Length 36;
Best Local Similarity 100.0%; Pred. No. 7.6e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      10 NDTDFVFLNNT 20
Db      17 NDTDFVFLNNT 27

RESULT 8
US-09-350-369D-1703
; Sequence 1703, Application US/09350369D
; GENERAL INFORMATION:
; APPLICANT: Samuel E. Hopkins
; APPLICANT: Elizabeth M. Dimassimo
; APPLICANT: Pamela P. Ruesnak
; APPLICANT: Thomas Michael Venetta
; TITLE OF INVENTION: Methods and Compositions for Detection of HIV GP41 Peptides and H
; FILE REFERENCE: 7872-068-999
; CURRENT APPLICATION NUMBER: US/09/350,369D
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 2088
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1703
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus-1 (HIV-1)
US-09-350-369D-1703

Query Match      55.0%; Score 11; DB 5; Length 36;
Best Local Similarity 100.0%; Pred. No. 7.6e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      10 NDTDFVFLNNT 20
Db      17 NDTDFVFLNNT 27

RESULT 9
US-09-350-369E-1703
; Sequence 1703, Application US/09350369E
; GENERAL INFORMATION:
; APPLICANT: Samuel E. Hopkins
; APPLICANT: Elizabeth M. Dimassimo
; APPLICANT: Pamela P. Ruesnak
; APPLICANT: Thomas Michael Venetta
; TITLE OF INVENTION: Methods and Compositions for Detection of HIV GP41 Peptides and H
; FILE REFERENCE: 7872-068-999
; CURRENT APPLICATION NUMBER: US/09/350,369E
```

```
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 2088
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1703
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus-1 (HIV-1)
US-09-350-369E-1703

Query Match      55.0%; Score 11; DB 5; Length 36;
Best Local Similarity 100.0%; Pred. No. 7.6e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      10 NDTDFVFLNNT 20
Db      17 NDTDFVFLNNT 27

RESULT 10
US-10-351-641-1084
; Sequence 1084, Application US/10351641
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
; APPLICANT: Lambert, D.
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; TITLE OF INVENTION: PROPERTIES
; FILE REFERENCE: 7872-100
; CURRENT APPLICATION NUMBER: US/10/351,641
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: 09/350,641
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/315,304
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1757
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1084
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-10-351-641-1084

Query Match      55.0%; Score 11; DB 6; Length 36;
Best Local Similarity 100.0%; Pred. No. 7.6e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      10 NDTDFVFLNNT 20
Db      17 NDTDFVFLNNT 27

RESULT 11
US-10-296-734-814
; Sequence 814, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 814
; LENGTH: 2010
```

```
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette B
US-10-296-734-814

Query Match      55.0%; Score 11; DB 6; Length 2010;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      10 NDTDFVFLNNT 20
Db      1479 NDTDFVFLNNT 1489

RESULT 12
US-10-296-734-406
; Sequence 406, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a consensus polyprotein
US-10-296-734-406

Query Match      55.0%; Score 11; DB 6; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      10 NDTDFVFLNNT 20
Db      532 NDTDFVFLNNT 542

RESULT 13
US-10-296-734-810
; Sequence 810, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 810
; LENGTH: 5985
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a savine
US-10-296-734-810

Query Match      55.0%; Score 11; DB 6; Length 5985;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      10 NDTDFVFLNNT 20
```

```
Db      3489 NDTDFVFLNNT 3499

RESULT 14
US-10-440-390-38
; Sequence 38, Application US/10440390
; GENERAL INFORMATION:
; APPLICANT: Lauer et al.
; TITLE OF INVENTION: Epitopes of Hepatitis C Virus
; FILE REFERENCE: 24028-011
; CURRENT APPLICATION NUMBER: US/10/440,390
; CURRENT FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: 60/381,273
; PRIOR FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-440-390-38

Query Match      50.0%; Score 10; DB 6; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.00033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      8 GENDTDFVL 17
Db      1 GENDTDFVL 10

RESULT 15
US-10-651-165-37
; Sequence 37, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-37

Query Match      40.0%; Score 8; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.053;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GPTYSWG 8
Db      13 GPTYSWG 20

Search completed: November 21, 2003, 22:12:57
Job time : 9.55 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:46:27 ; Search time 9.25 Seconds
(without alignments)
207.932 Million cell updates/sec

Title: US-09-973-025-82

Perfect score: 20

Sequence: 1 GAPTYSWGENDTDVFLNNT 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR 76: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	3011	1 GNVVC3	genome polyprotein
2	11	55.0	640	2 JQ1584	genome polyprotein
3	11	55.0	716	2 JQ1366	polyprotein - hepa
4	11	55.0	3011	1 GNVVCH	genome polyprotein
5	11	55.0	3011	1 S40770	genome polyprotein
6	10	50.0	3010	1 GNVVCJ	genome polyprotein
7	10	50.0	3010	1 GNVVTJ	genome polyprotein
8	8	40.0	235	2 S32747	genome polyprotein
9	6	30.0	172	2 D87584	RNA polymerase sig
10	6	30.0	234	2 S32742	genome polyprotein
11	6	30.0	287	2 AF2333	hypothetical prote
12	6	30.0	300	2 A97329	transcription regu
13	6	30.0	313	2 G81173	adhesin mafa - Nei
14	6	30.0	313	2 G81782	adhesin mafa2 NMA2
15	6	30.0	318	2 T08845	cysteine proteinas
16	6	30.0	579	2 D90058	hypothetical prote
17	6	30.0	638	2 S22490	acetolactate synth
18	6	30.0	638	2 S22491	acetolactate synth
19	6	30.0	734	2 T37696	hypothetical prote
20	6	30.0	787	2 PN0677	hypothetical prote
21	6	30.0	809	2 G86603	uramate-Ala ligas
22	6	30.0	809	2 D72022	UDP-N-acetylmuram
23	6	30.0	871	2 S47518	cadherin - African
24	6	30.0	1648	2 S57163	probable membrane
25	6	30.0	1657	2 T25421	hypothetical prote
26	5	25.0	20	2 A47687	3-dehydroquinat d
27	5	25.0	53	2 G82546	hypothetical prote
28	5	25.0	58	2 G36491	phosphoprotein pho
29	5	25.0	58	2 E36491	phosphoprotein pho

30 5 25.0 67 2 H95273 probable CspA5 col
31 5 25.0 69 2 AC2940 cold shock protein
32 5 25.0 85 2 T18137 hypothetical prote
33 5 25.0 90 2 T17977 hypothetical prote
34 5 25.0 92 2 D64029 hypothetical prote
35 5 25.0 93 2 S31185 outer membrane pro
36 5 25.0 93 2 S33842 acyl-CoA thioester
37 5 25.0 95 2 JQ1483 hypothetical 11.2K
38 5 25.0 95 2 F83887 hypothetical prote
39 5 25.0 96 2 B36074 cysteine proteinas
40 5 25.0 98 2 E98342 probable cold shoc
41 5 25.0 100 2 H82800 hypothetical prote
42 5 25.0 101 2 AH3316 cold shock protein
43 5 25.0 102 2 H69187 hypothetical prote
44 5 25.0 112 2 D87415 nitrogen regulator
45 05 25.0 113 2 I73629 phosphoprotein pho

ALIGNMENTS

RESULT 1

GNVVC3

genome polyprotein - hepatitis C virus (strain HCV-1)
N:Contains: capsid protein C; envelope protein M; hepatitisvirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C>Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 19-Jan-2001
C:Accession: A39166; PQ0403; PQ0404
R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Co
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A:Title: Genetic organization and diversity of the hepatitis C virus.
A:Reference number: A39166; MUID:91172826; PMID:1848704
A:Accession: A39166
A:Molecule type: mRNA
A:Residues: 1-3011 <CHO>
A:Cross-references: GB:M62321; NID:G329873; PIDN:AAA45676.1; PID:G329874
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.
J. Gen. Virol. 73, 1131-1141, 1992
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to
A:Reference number: PQ0393; MUID:92268871; PMID:1316933
A:Accession: PQ0403
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CHA>
A:Cross-references: DBJ:D10128
A:Experimental source: isolates E-b16
A:Accession: PQ0404
A>Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CH2>
A:Experimental source: isolates E-b17
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <BPM>
F:192-389/Product: major envelope protein E #status predicted <MEB>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitisvirin #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 100.0% Score 20; DB 1; Length 3011;

Best Local Similarity 100.0%; Pred. No. 6.6e-14; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGENDTDVFLNNT 20

|||||

Db 523 GAPTYSWGENDTDFVLNNT 542

RESULT 2

JQ1584

genome polyprotein - hepatitis C virus (strain U.K.) (fragment)

N:Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural protein NS4a; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000

C:Accession: JQ1584

R:Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.

J. Gen. Virol. 73, 1521-1525, 1992

A:Title: Cloning and sequencing of the structural region and expression of putative core protein E2 and nonstructural protein NS5

A:Reference number: JQ1584; MUID:92300349; PMID:1318944

A:Accession: JQ1584

A:Molecule type: genomic RNA

A:Residues: 1-640 <KUN>

A:Cross-references: GB:X84079; NID:G643119; PIDN:CAA58888.1; PID:G643120

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; polyprotein

F:1-191/Product: core protein C #status predicted <CPC>

F:192-389/Product: envelope protein E1 #status predicted <EE1>

F:390-640/Product: envelope protein E2 and nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn) (coverage)

Query Match 55.0%; Score 11; DB 2; Length 640;

Best Local Similarity 100.0%; Pred. No. 0.00015; Mismatches 0; Indels 0; Gaps 0;

Matches 11; Conservative 0;

Qy 10 NDTDFVLNNT 20

Db 532 NDTDFVLNNT 542

RESULT 3

JQ1366

polyprotein - hepatitis C virus (French isolate) (fragments)

C:Species: hepatitis C virus

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000

C:Accession: JQ1366

R:Kremsdorff, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.

J. Gen. Virol. 72, 2557-2561, 1991

A:Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication for the origin of the virus

A:Reference number: JQ1366; MUID:92013977; PMID:1655961

A:Accession: JQ1366

A:Molecule type: genomic RNA

A:Residues: 1-716 <KRE>

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: glycoprotein; polyprotein

F:84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #status predicted <NS5>

Query Match 55.0%; Score 11; DB 2; Length 716;

Best Local Similarity 100.0%; Pred. No. 0.00016; Mismatches 0; Indels 0; Gaps 0;

Matches 11; Conservative 0;

Qy 10 NDTDFVLNNT 20

Db 199 NDTDFVLNNT 209

RESULT 4

GNVCH

genome polyprotein - hepatitis C virus (strain H)

N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein NS4a; nonstructural protein NS5b; nonstructural protein NS5

C:Species: hepatitis C virus

A:Note: host Homo sapiens (man)

C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001

C:Accession: A36814; A41546

R:Inchauspe, G.; Zebede, S.; Lee, D.H.; Sugitani, M.; Prince, A.M.

submitted to GenBank, July 1992

A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: comparison with other strains

A:Reference number: A36814

A:Accession: A36814

A:Molecule type: genomic RNA

A:Residues: 1-3011 <INC>

A:Cross-references: GB:M67463; NID:G329737; PIDN:AAA45534.1; PID:G329738

R:Inchauspe, G.; Zebede, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.

Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991

A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: comparison with other strains

A:Reference number: A41546; MUID:92052256; PMID:1659800

A:Contents: annotation

A:Note: neither amino acid nor nucleotide sequence is given

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein C; envelope protein; glycoprotein; hydrolase; nonstructural protein NS4a; nonstructural protein NS5b; nonstructural protein NS5

F:1-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitis C virus genome polyprotein NS5 #status predicted <NS5>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4A>

F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,2333,2334,2335,2336,2337,2338,2339,2340,2341,2342,2343,2344,2345,2346,2347,2348,2349,2350,2351,2352,2353,2354,2355,2356,2357,2358,2359,2360,2361,2362,2363,2364,2365,2366,2367,2368,2369,2370,2371,2372,2373,2374,2375,2376,2377,2378,2379,2380,2381,2382,2383,2384,2385,2386,2387,2388,2389,2390,2391,2392,2393,2394,2395,2396,2397,2398,2399,2400,2401,2402,2403,2404,2405,2406,2407,2408,2409,2410,2411,2412,2413,2414,2415,2416,2417,2418,2419,2420,2421,2422,2423,2424,2425,2426,2427,2428,2429,2430,2431,2432,2433,2434,2435,2436,2437,2438,2439,2440,2441,2442,2443,2444,2445,2446,2447,2448,2449,2450,2451,2452,2453,2454,2455,2456,2457,2458,2459,2460,2461,2462,2463,2464,2465,2466,2467,2468,2469,2470,2471,2472,2473,2474,2475,2476,2477,2478,2479,2480,2481,2482,2483,2484,2485,2486,2487,2488,2489,2490,2491,2492,2493,2494,2495,2496,2497,2498,2499,2500,2501,2502,2503,2504,2505,2506,2507,2508,2509,2510,2511,2512,2513,2514,2515,2516,2517,2518,2519,2520,2521,2522,2523,2524,2525,2526,2527,2528,2529,2530,2531,2532,2533,2534,2535,2536,2537,2538,2539,2540,2541,2542,2543,2544,2545,2546,2547,2548,2549,2550,2551,2552,2553,2554,2555,2556,2557,2558,2559,2560,2561,2562,2563,2564,2565,2566,2567,2568,2569,2570,2571,2572,2573,2574,2575,2576,2577,2578,2579,2580,2581,2582,2583,2584,2585,2586,2587,2588,2589,2590,2591,2592,2593,2594,2595,2596,2597,2598,2599,2600,2601,2602,2603,2604,2605,2606,2607,2608,2609,2610,2611,2612,2613,2614,2615,2616,2617,2618,2619,2620,2621,2622,2623,2624,2625,2626,2627,2628,2629,2630,2631,2632,2633,2634,2635,2636,2637,2638,2639,2640,2641,2642,2643,2644,2645,2646,2647,2648,2649,2650,2651,2652,2653,2654,2655,2656,2657,2658,2659,2660,2661,2662,2663,2664,2665,2666,2667,2668,2669,2670,2671,2672,2673,2674,2675,2676,2677,2678,2679,2680,2681,2682,2683,2684,2685,2686,2687,2688,2689,2690,2691,2692,2693,2694,2695,2696,2697,2698,2699,2700,2701,2702,2703,2704,2705,2706,2707,2708,2709,2710,2711,2712,2713,2714,2715,2716,2717,2718,2719,2720,2721,2722,2723,2724,2725,2726,2727,2728,2729,2730,2731,2732,2733,2734,2735,2736,2737,2738,2739,2740,2741,2742,2743,2744,2745,2746,2747,2748,2749,2750,2751,2752,2753,2754,2755,2756,2757,2758,2759,2760,2761,2762,2763,2764,2765,2766,2767,2768,2769,2770,2771,2772,2773,2774,2775,2776,2777,2778,2779,2780,2781,2782,2783,2784,2785,2786,2787,2788,2789,2790,2791,2792,2793,2794,2795,2796,2797,2798,2799,2800,2801,2802,2803,2804,2805,2806,2807,2808,2809,2810,2811,2812,2813,2814,2815,2816,2817,2818,2819,2820,2821,2822,2823,2824,2825,2826,2827,2828,2829,2830,2831,2832,2833,2834,2835,2836,2837,2838,2839,2840,2841,2842,2843,2844,2845,2846,2847,2848,2849,2850,2851,2852,2853,2854,2855,2856,2857,2858,2859,2860,2861,2862,2863,2864,2865,2866,2867,2868,2869,2870,2871,2872,2873,2874,2875,2876,2877,2878,2879,2880,2881,2882,2883,2884,2885,2886,2887,2888,2889,2890,2891,2892,2893,2894,2895,2896,2897,2898,2899,2900,2901,2902,2903,2904,2905,2906,2907,2908,2909,2910,2911,2912,2913,2914,2915,2916,2917,2918,2919,2920,2921,2922,2923,2924,2925,2926,2927,2928,2929,2930,2931,2932,2933,2934,2935,2936,2937,2938,2939,2940,2941,2942,2943,2944,2945,2946,2947,2948,2949,2950,2951,2952,2953,2954,2955,2956,2957,2958,2959,2960,2961,2962,2963,2964,2965,2966,2967,2968,2969,2970,2971,2972,2973,2974,2975,2976,2977,2978,2979,2980,2981,2982,2983,2984,2985,2986,2987,2988,2989,2990,2991,2992,2993,2994,2995,2996,2997,2998,2999,3000,3001,3002,3003,3004,3005,3006,3007,3008,3009,3010,3011,3012,3013,3014,3015,3016,3017,3018,3019,3020,3021,3022,3023,3024,3025,3026,3027,3028,3029,3030,3031,3032,3033,3034,3035,3036,3037,3038,3039,3040,3041,3042,3043,3044,3045,3046,3047,3048,3049,3050,3051,3052,3053,3054,3055,3056,3057,3058,3059,3060,3061,3062,3063,3064,3065,3066,3067,3068,3069,3070,3071,3072,3073,3074,3075,3076,3077,3078,3079,3080,3081,3082,3083,3084,3085,3086,3087,3088,3089,3090,3091,3092,3093,3094,3095,3096,3097,3098,3099,3100,3101,3102,3103,3104,3105,3106,3107,3108,3109,3110,3111,3112,3113,3114,3115,3116,3117,3118,3119,3120,3121,3122,3123,3124,3125,3126,3127,3128,3129,3130,3131,3132,3133,3134,3135,3136,3137,3138,3139,3140,3141,3142,3143,3144,3145,3146,3147,3148,3149,3150,3151,3152,3153,3154,3155,3156,3157,3158,3159,3160,3161,3162,3163,3164,3165,3166,3167,3168,3169,3170,3171,3172,3173,3174,3175,3176,3177,3178,3179,3180,3181,3182,3183,3184,3185,3186,3187,3188,3189,3190,3191,3192,3193,3194,3195,3196,3197,3198,3199,3200,3201,3202,3203,3204,3205,3206,3207,3208,3209,3210,3211,3212,3213,3214,3215,3216,3217,3218,3219,3220,3221,3222,3223,3224,3225,3226,3227,3228,3229,3230,3231,3232,3233,3234,3235,3236,3237,3238,3239,3240,3241,3242,3243,3244,3245,3246,3247,3248,3249,3250,3251,3252,3253,3254,3255,3256,3257,3258,3259,3260,3261,3262,3263,3264,3265,3266,3267,3268,3269,3270,3271,3272,3273,3274,3275,3276,3277,3278,3279,3280,3281,3282,3283,3284,3285,3286,3287,3288,3289,3290,3291,3292,3293,3294,3295,3296,3297,3298,3299,3300,3301,3302,3303,3304,3305,3306,3307,3308,3309,3310,3311,3312,3313,3314,3315,3316,3317,3318,3319,3320,3321,3322,3323,3324,3325,3326,3327,3328,3329,3330,3331,3332,3333,3334,3335,3336,3337,3338,3339,3340,3341,3342,3343,3344,3345,3346,3347,3348,3349,3350,3351,3352,3353,3354,3355,3356,3357,3358,3359,3360,3361,3362,3363,3364,3365,3366,3367,3368,3369,3370,3371,3372,3373,3374,3375,3376,3377,3378,3379,3380,3381,3382,3383,3384,3385,3386,3387,3388,3389,3390,3391,3392,3393,3394,3395,3396,3397,3398,3399,3400,3401,3402,3403,3404,3405,3406,3407,3408,3409,3410,3411,3412,3413,3414,3415,3416,3417,3418,3419,3420,3421,3422,3423,3424,3425,3426,3427,3428,3429,3430,3431,3432,3433,3434,3435,3436,3437,3438,3439,3440,3441,3442,3443,3444,3445,3446,3447,3448,3449,3450,3451,3452,3453,3454,3455,3456,3457,3458,3459,3460,3461,3462,3463,3464,3465,3466,3467,3468,3469,3470,3471,3472,3473,3474,3475,3476,3477,3478,3479,3480,3481,3482,3483,3484,3485,3486,3487,3488,3489,3490,3491,3492,3493,3494,3495,3496,3497,3498,3499,3500,3501,3502,3503,3504,3505,3506,3507,3508,3509,3510,3511,3512,3513,3514,3515,3516,3517,3518,3519,3520,3521,3522,3523,3524,3525,3526,3527,3528,3529,3530,3531,3532,3533,3534,3535,3536,3537,3538,3539,3540,3541,3542,3543,3544,3545,3546,3547,3548,3549,3550,3551,3552,3553,3554,3555,3556,3557,3558,3559,3560,3561,3562,3563,3564,3565,3566,3567,3568,3569,3570,3571,3572,3573,3574,3575,3576,3577,3578,3579,3580,3581,3582,3583,3584,3585,3586,3587,3588,3589,3590,3591,3592,3593,3594,3595,3596,3597,3598,3599,3600,3601,3602,3603,3604,3605,3606,3607,3608,3609,3610,3611,3612,3613,3614,3615,3616,3617,3618,3619,3620,3621,3622,3623,3624,3625,3626,3627,3628,3629,3630,3631,3632,3633,3634,3635,3636,3637,3638,3639,3640,3641,3642,3643,3644,3645,3646,3647,3648,3649,3650,3651,3652,3653,3654,3655,3656,3657,3658,3659,3660,3661,3662,3663,3664,3665,3666,3667,3668,3669,3670,3671,3672,3673,3674,3675,3676,3677,3678,3679,3680,3681,3682,3683,3684,3685,3686,3687,3688,3689,3690,3691,3692,3693,3694,3695,3696,3697,3698,3699,3700,3701,3702,3703,3704,3705,3706,3707,3708,3709,3710,3711,3712,3713,3714,3715,3716,3717,3718,3719,3720,3721,3722,3723,3724,3725,3726,3727,3728,3729,3730,3731,3732,3733,3734,3735,3736,3737,3738,3739,3740,3741,3742,3743,3744,3745,3746,3747,3748,3749,3750,3751,3752,3753,3754,3755,3756,3757,3758,3759,3760,3761,3762,3763,3764,3765,3766,3767,3768,3769,3770,3771,3772,3773,3774,3775,3776,3777,3778,3779,3780,3781,3782,3783,3784,3785,3786,3787,3788,3789,3790,3791,3792,3793,3794,3795,3796,3797,3798,3799,3800,3801,3802,3803,3804,3805,3806,3807,3808,3809,3810,3811,3812,3813,3814,3815,3816,3817,3818,3819,3820,3821,3822,3823,3824,3825,3826,3827,3828,3829,3830,3831,3832,3833,3834,3835,3836,3837,3838,3839,3840,3841,3842,3843,3844,3845,3846,3847,3848,3849,3850,3851,3852,3853,3854,3855,3856,3857,3858,3859,3860,3861,3862,3863,3864,3865,3866,3867,3868,3869,3870,3871,3872,3873,3874,3875,3876,3877,3878,3879,3880,3881,3882,3883,3884,3885,3886,3887,3888,3889,3890,3891,3892,3893,3894,3895,3896,3897,3898,3899,3900,3901,3902,3903,3904,3905,3906,3907,3908,3909,3910,3911,3912,3913,3914,3915,3916,3917,3918,3919,3920,3921,3922,3923,3924,3925,3926,3927,3928,3929,3930,3931,3932,3933,3934,3935,3936,3937,3938,3939,3940,3941,3942,3943,3944,3945,3946,3947,3948,3949,3950,3951,3952,3953,3954,3955,3956,3957,3958,3959,3960,3961,3962,3963,3964,3965,3966,3967,3968,3969,3970,3971,3972,3973,3974,3975,3976,3977,3978,3979,3980,3981,3982,3983,3984,3985,3986,3987,3988,3989,3990,3991,3992,3993,3994,3995,3996,3997,3998,3999,4000,4001,4002,4003,4004,4005,4006,4007,4008,4009,4010,4011,4012,4013,4014,4015,4016,4017,4018,4019,4020,4021,4022,4023,4024,4025,4026,4027,4028,4029,4030,4031,4032,4033,4034,4035,4036,4037,4038,4039,4040,4041,4042,4043,4044,4045,4046,4047,4048,4049,4050,4051,4052,4053,4054,4055,4056,4057,4058,4059,4060,4061,4062,4063,4064,4065,4066,4067,4068,4069,4070,4071,4072,4073,4074,4075,4076,4077,4078,4079,4080,4081,4082,4083,4084,4085,4086,4087,4088,4089,4090,4091,4092,4093,4094,4095,4096,4097,4098,4099,4100,4101,4102,4103,4104,4105,4106,4107,4108,4109,4110,4111,4112,4113,4114,4115,4116,4117,4118,4119,4120,4121,4122,4123,4124,4125,4126,4127,4128,4129,4130,4131,4132,4133,4134,4135,4136,4137,4138,4139,4140,4141,4142,4143,4144,4145,4146,4147,4148,4149,4150,4151,4152,4153,4154,4155,4156,4157,4158,4159,4160,4161,4162,4163,4164,4165,4166,4167,4168,4169,4170,4171,4172,4173,4174,4175,4176,4177,4178,4179,4180,4181,4182,4183,4184,4185,4186,4187,4188,4189,4190,4191,4192,4193,4194,4195,4196,4197,4198,4199,4200,4201,4202,4203,4204,4205,4206,4207,4208,4209,4210,4211,4212,4213,4214,4215,4216,4217,4218,4219,4220,4221,4222,4223,4224,4225,4226,4227,4228,4229,4230,4231,4232,4233,4234,4235,4236,4237,4238,4239,4240,4241,4242,4243,4244,4245,4246,4247,4248,4249,4250,4251,4252,4253,4254,4255,4256,4257,4258,4259,4260,4261,4262,4263,4264,4265,4266,4267,4268,4269,4270,4271,4272,4273,4274,4275,4276,4277,4278,4279,4280,4281,4282,4283,4284,4285,4286,4287,4288,4289,4290,4291,4292,4293,4294,4295,4296,4297,4298,4299,4300,4301,4302,4303,4304,4305,4306,4307,4308,4309,4310,4311,4312,4313,4314,4315,4316,4317,4318,4319,4320,4321,4322,4323,4324,4325,4326,4327,4328,4329,4330,4331,4332,4333,4334,4

Query Match 55.0%; Score 11; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 0.00061;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 NDTDFVFLNNT 20
 |||||
 DB 532 NDTDFVFLNNT 542

RESULT 6
 GNMVUJ
 genome polypotein - hepatitis C virus (strain J)
 N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
 C:Accession: A39253; PS0086
 R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimoto Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
 A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients
 A:Reference number: A39253; MUID:91088550; PMID:2175903
 A:Accession: A39253
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <KAT>
 A:Cross-references: GB:D90208; NID:G221610; PIDN:BAAL4233.1; PID:G221611
 R:Kato, N.; Ohkoshi, S.; Shimotohno, K.
 Proc. Jpn. Acad. 65B, 219-223, 1989
 A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari
 A:Reference number: PS0085
 A:Accession: PS0086
 A:Molecule type: genomic RNA
 A:Residues: 2650-2707 <KA2>
 A:Experimental source: Japanese isolate
 C:Comment: The cleavage sites of this polypotein have not been determined.
 C:Superfamily: hepatitis C virus genome polypotein
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypotein; serin
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: capsid protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEP>
 F:390-729/Product: major envelope protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

Query Match 50.0%; Score 10; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.0078;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGEN 10
 |||||
 DB 523 GAPTYSWGEN 532

RESULT 7
 GNMVTV
 genome polypotein - hepatitis C virus (strain Taiwan)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain Taiwan)
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 A:Note: host Homo sapiens (man)
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C:Accession: A40244
 R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
 Virology 188, 102-113, 1992
 A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
 A:Reference number: A40244; MUID:92230206; PMID:1314449

A:Accession: A40244
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <CHE>
 A:Cross-references: GB:M84754
 C:Superfamily: hepatitis C virus genome polypotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: capsid protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEP>
 F:390-729/Product: major envelope protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,2240,2

Query Match 50.0%; Score 10; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.0078;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAPTYSWGEN 10
 |||||
 DB 523 GAPTYSWGEN 532

RESULT 8
 S32747
 genome polypotein - hepatitis C virus (isolate HU-1) (fragment)
 N:Contains: envelope protein E2
 C:Species: hepatitis C virus
 A:Variety: isolate HU-1
 C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S32747
 R:Rogendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
 submitted to the EMBL Data Library, February 1993
 A:Description: Variability of the envelope regions of HCV in European isolates and its
 A:Reference number: S32741
 A:Accession: S32747
 A:Molecule type: genomic RNA
 A:Residues: 1-235 <ROG>
 A:Cross-references: EMBL:X72977; NID:G296112; PIDN:CAA51482.1; PID:G296113
 A:Experimental source: isolate HU-1
 C:Superfamily: hepatitis C virus genome polypotein
 C:Keywords: envelope protein; polypotein
 F:1-235/Product: envelope protein E2 #status predicted <MAT>

Query Match 40.0%; Score 8; DB 2; Length 235;
 Best Local Similarity 100.0%; Pred. No. 0.12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PTYSWGEN 10
 |||||
 DB 157 PTYSWGEN 164

RESULT 9
 D87584
 RNA polymerase sigma-70 factor, ECF subfamily [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: D87584
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; deBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 n, J.; Emdolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; MUID:21173698; PMID:11259647
 A:Accession: D87584
 A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-172 <STO>
A:Cross-references: GB:AE005673; NID:gl13424294; PIDN:AAK24672.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2707

Query Match 30.0%; Score 6; DB 2; Length 172;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 DVFVLN 18
|||||
DB 129 DVFVLN 134

RESULT 10
S32742
genome polyprotein - hepatitis C virus (isolate CR-1) (fragment)
N:Contains: envelope protein E2
C:Species: hepatitis C virus
A:Variety: isolate CR-1
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S32742
R:Rogendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
submitted to the EMBL Data Library, February 1993
A:Description: Variability of the envelope regions of HCV in European isolates and its effect on the virus genome
A:Reference number: S32741
A:Accession: S32742
A:Molecule type: genomic RNA
A:Residues: 1-234 <ROG>
A:Cross-references: EMBL:X72979; NID:g296102; PIDN:CAA51484.1; PID:g296103
A:Experimental source: isolate CR-1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; polyprotein
F:1-234/Product: envelope protein E2 #status predicted <MAT>

Query Match 30.0%; Score 6; DB 2; Length 234;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PTYSWG 8
|||||
DB 156 PTYSWG 161

RESULT 11
AF2333
hypothetical protein all4221 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 06-Jan-2003
C:Accession: AF2333
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF2333
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-287 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAW75920.1; PID:gl17133356; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
C:Superfamily: tropinesterase

Query Match 30.0%; Score 6; DB 2; Length 287;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 WGENDT 12
|||||

Db 239 WGENDT 240

RESULT 12
A97329
transcription regulator, YORV B. subtilis homolog [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: A97329
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: A97329
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-300 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK81420.1; PID:gl15026585; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CRC3494

Query Match 30.0%; Score 6; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GENDTD 13
|||||
DB 109 GENDTD 114

RESULT 13
GB1173
adhesin mafa - Neisseria meningitidis (strain MC58 serogroup B)
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: GB1173
R:Tetelin, M.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; H. H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizzi, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755; PMID:10710307
A:Accession: GB1173
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-313 <TET>
A:Cross-references: GB:AE002420; GB:AE002098; NID:g7225876; PIDN:AAFA1073.1; PID:g722588
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0652
C:Superfamily: Neisseria adhesin mafa

Query Match 30.0%; Score 6; DB 2; Length 313;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 DTDVFP 16
|||||
DB 190 DTDVFP 195

RESULT 14
GB1782
adhesin MafA2 NMA2112 [imported] - Neisseria meningitidis (strain Z2491 serogroup A)
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: GB1782
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jørgensen, K.; Leather, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: G81782
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-313 <PAR>
A:Cross-references: GB:AL162758; GB:AL157959; NID:G7380672; PIDN:CAB85326.1; PID:G738073
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: mafa2; NMA2112
C:Superfamily: *Neisseria* adhesin mafa

Query Match 30.0%; Score 6; DB 2; Length 313;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 DTDVFEV 16
Db 190 DTDVFEV 195

RESULT 15

T08845
cysteine proteinase (EC 3.4.22.-) isoform A - soybean (fragment)
N:Alternate names: thiol proteinase isoform A
C:Species: Glycine max (soybean)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 23-Jul-1999
C:Accession: T08845
R:Herman, E.M.; Kalinski, A.J.; Rowley, D.
submitted to the EMBL Data Library, September 1996
A:Description: Soybean leaf thiol protease, isoform A.
A:Reference number: Z16489
A:Accession: T08845
A:Status: translated from GB/EMBL/DBBJ
A:Molecule type: mRNA
A:Residues: 1-318 <HER>
A:Cross-references: EMBL:U71380; NID:G1619904; PID:G1619905
A:Experimental source: cultivar Century; leaf
C:Superfamily: papain
C:Keywords: cysteine proteinase; hydrolase
F:109-143/Disulfide bonds: #status predicted
F:112,252,279/Active site: Cys, His, Asn #status predicted
F:191-246/Disulfide bonds: #status predicted
F:279,280,281/Binding site: substrate (Asn, Ser, Trp) #status predicted

Query Match 30.0%; Score 6; DB 2; Length 318;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 SWGEND 11
Db 284 SWGEND 289

Search completed: November 21, 2003, 21:11:33
Job time : 10.25 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:52 ; Search time 4.85 Seconds
(without alignments)
193.925 Million cell updates/sec

Title: US-09-973-025-82

Perfect score: 20

Sequence: 1 GAPTYSWGENDTDFVLNNT 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	3011	1 POLG_HCV1	P26664 h genome po
2	11	55.0	3011	1 POLG_HCVH	P27958 h genome po
3	10	50.0	3010	1 POLG_HCVJA	P26662 h genome po
4	10	50.0	3010	1 POLG_HCVTW	P29846 h genome po
5	8	40.0	737	1 POLG_HCVJ7	P27961 hepatitis c
6	6	30.0	452	1 GUX1_CRYPA	Q00548 cryphonectr
7	6	30.0	734	1 YDZB_SCHPO	O13718 schizosach
8	6	30.0	773	1 PNK4_HUMAN	Q9NVE7 homo sapien
9	6	30.0	773	1 PNK4_RAT	Q92388 rattus norv
10	6	30.0	809	1 MUDJ_CHLPN	Q9Z701 c murc/ddl
11	6	30.0	872	1 CAD1_XENLA	P30944 xenopus lae
12	6	30.0	946	1 AMYG_CANAL	O74254 candida alb
13	6	30.0	1648	1 YJ9H_YEAST	P47171 saccharomyc
14	5	25.0	20	1 AROO_AMEYE	P46380 amycolatops
15	5	25.0	66	1 CSP2_LACPL	P96349 lactobacill
16	5	25.0	69	1 CSPA_RHIME	Q9Z386 rhizobium m
17	5	25.0	92	1 YE21_HAEIN	P44192 haemophilus
18	5	25.0	96	1 PAP5_CARPA	P05593 carica papa
19	5	25.0	98	1 FER_THANE	Q98450 thalassios
20	5	25.0	114	1 YIE5_STRCO	Q9fc38 streptomyce
21	5	25.0	120	1 VE4_HPV42	P27225 human papil
22	5	25.0	131	1 RS8E_PYRAE	Q8Ztb8 pyrobaculum
23	5	25.0	131	1 YN16_YEAST	P53837 saccharomyc
24	5	25.0	144	1 AROO_RHILO	Q98nc1 rhizobium l
25	5	25.0	157	1 AROO_BRUME	Q8Y9u6 bruceella me
26	5	25.0	181	1 EPID_STAEP	P30197 staphylococ
27	5	25.0	182	1 WCAF_ECOLI	P71240 escherichia
28	5	25.0	206	1 GTS6_CAREL	P91252 caenorhabdi
29	5	25.0	206	1 RNFG_SALTI	Q8Z6q7 salmonella
30	5	25.0	206	1 RNFG_SALTY	Q8Zpm4 salmonella
31	5	25.0	213	1 NEUN_CARAU	P17691 carassius a
32	5	25.0	218	1 IM23_SCHHA	Q26499 schistosoma
33	5	25.0	218	1 IM23_SCHMA	P19331 schistosoma

34 5 25.0 219 1 VB02_VACCC
35 5 25.0 219 1 VB02_VACCV
36 5 25.0 220 1 PCP_THELI
37 5 25.0 226 1 VATE_MESCR
38 5 25.0 227 1 SRPB_SCHPO
39 5 25.0 228 1 ENGB_BUCBP
40 5 25.0 229 1 VATE_SPIOL
41 5 25.0 231 1 YG65_ARCFU
42 5 25.0 234 1 IPT_PSESS
43 5 25.0 246 1 CAH3_CABEL
44 5 25.0 248 1 SCTM_HUMAN
45 5 25.0 253 1 NAPI_MOUSE

P20999 vaccinia vi
Q01225 vaccinia vi
O07883 thermococu
Q40272 mesembryant
O13950 schizosach
Q89ad0 buchnera ap
Q41396 spinacia ol
Q23003 archaeoglob
P06619 pseudomonas
Q27504 caenorhabdi
Q89vn6 homo sapien
Q61200 mus musculus

ALIGNMENTS

RESULT 1
POLG_HCV1
ID POLG_HCV1 STANDARD; PRT; 3011 AA.
AC P26664.0
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate 1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA Bradley D.W., Kuo G., Houghton M.;
RT "Genetic organization and diversity of the hepatitis C virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA)(N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: OF
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M62321; AAA45676.1; -.
CC PIR; A39166; GNMVVC3.
CC PDB; 1A1V; 16-FEB-99.
CC PDB; 1HEI; 25-NOV-98.
CC MEROPS; S29.001; -.
CC MEROPS; U39.001; -.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002522; HCV_capsid.

CC DR EMBL; MG7463; AAA4534.1; -
DR PIR; A36814; GNVVCH.
DR PDB; 1HEI; 25-NOV-98.
DR PDB; 1AIV; 16-FEB-99.
DR PDB; 1AIR; 17-JUN-98.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR TRANSFAC; T04155; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA pol DS PS.
DR InterPro; IPR007094; RNA pol PS vir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase C; 1.
DR Pfam; PF00998; Viral RdRp; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC7; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
FT CHAIN 192 383 CAPSID PROTEIN C.
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
FT CHAIN 810 1026 PROTEIN P7.
FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5A.
FT CHAIN 3011 3011 NONSTRUCTURAL PROTEIN NS5B.
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DESCH BOX.
FT CARBOHYD 136 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).

FT STRAND 1224 1226
FT TURN 1232 1233
FT TURN 1236 1238
FT HELIX 1239 1246
FT TURN 1247 1248
FT STRAND 1251 1255
FT HELIX 1258 1271
FT TURN 1272 1272
FT STRAND 1277 1280
FT TURN 1281 1282
FT STRAND 1283 1285
FT STRAND 1291 1295
FT HELIX 1296 1301
FT TURN 1302 1303
FT STRAND 1312 1316
FT TURN 1317 1319
FT HELIX 1323 1335
FT TURN 1336 1340
FT STRAND 1343 1347
FT TURN 1352 1353
FT TURN 1360 1361
FT STRAND 1362 1366
FT STRAND 1368 1368
FT TURN 1376 1377
FT STRAND 1378 1380
FT HELIX 1382 1385
FT STRAND 1389 1393
FT HELIX 1397 1409
FT TURN 1410 1411
FT STRAND 1414 1417
FT TURN 1419 1420
FT STRAND 1432 1436
FT TURN 1438 1439
FT STRAND 1450 1453
FT STRAND 1456 1463
FT STRAND 1471 1478
FT STRAND 1480 1480
FT HELIX 1481 1488
FT TURN 1489 1490
FT STRAND 1497 1501
FT STRAND 1507 1507
FT STRAND 1511 1511
FT HELIX 1514 1527
FT STRAND 1532 1544
FT STRAND 1550 1550
FT HELIX 1555 1564
FT TURN 1570 1578
FT TURN 1579 1580
FT HELIX 1584 1597
FT TURN 1598 1598
FT HELIX 1606 1611
FT TURN 1614 1618
FT STRAND 1622 1623
FT STRAND 1627 1627
FT STRAND 1635 1636
FT HELIX 1640 1652
SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;
Query Match 55.0%; Score 11; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.00066;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 10 NDTQVFLNNT 20
Db 532 NDTQVFLNNT 542
RESULT 3
POLG_HCVJA 0
ID POLG_HCVJA STANDARD; PRT; 3010 AA.
AC P26662;
DT 01-AUG-1992 (Rel. 23, Created)

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230206; PubMed=1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 RT "The Taiwanese hepatitis C virus genome: sequence determination and
 mapping the 5' terminus of viral genomic and antigenomic RNA.";
 RL Virology 188:102-113(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA) (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/commence/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M84754; -; NOT_ANNOTATED_CDS.
 DR PIR; A40244; GNMVTV.
 DR PDB; 1N64; 25-FEB-03.
 DR PDB; INS3; 08-APR-98.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRp.
 DR InterPro; IPR007095; RNA pol DS PS.
 DR InterPro; IPR007094; RNA pol psvir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383

FT CHAIN 384 729
 FT CHAIN 1007 1066
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2529 2529
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;
 Query Match 50.0%; Score 10; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.0075;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GAPTYSWGEN 10
 DB 523 GAPTYSWGEN 532
 RESULT 5
 POLG_HCVJ7 STANDARD; PRT; 737 AA.
 ID POLG_HCVJ7
 AC P27961;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
 DE protein (Envelope protein M); Major envelope protein E; Nonstructural
 DE proteins NS1 and NS2] (fragment).
 OS Hepatitis C virus (isolate HC-J7) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11114;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tenda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Murgall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallard J., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Useery D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -!- SIMILARITY: TO YEAST YJL012C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z98596; CAB11204.1; -.
DR PIR; T37696; T37696.
DR GenesDB SPombe; SPAC14C4.11; -.
DR InterPro; IPR003807; DUF202.
DR InterPro; IPR004331; SPX.
DR Pfam; PF02656; DUF202; 1.
DR Pfam; PF03105; SPX; 1.
DR KW Hypothetical protein; Transmembrane.
FT TRANSMEM 470 490 POTENTIAL.
FT TRANSMEM 625 645 POTENTIAL.
FT TRANSMEM 651 671 POTENTIAL.
FT TRANSMEM 694 714 POTENTIAL.
SQ SEQUENCE 734 AA; 85445 MW; 2373306038C7F4EE CRC64;
Query Match 30.0%; Score 6; DB 1; Length 734;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 SWGEND 11
Db 37 SWGEND 42
RESULT 8
ID PNK4 HUMAN STANDARD; PRT; 773 AA.
AC Q9NV67; Q9H3X5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pantothenate kinase 4 (EC 2.7.1.33) (Pantothenic acid kinase 4)
DE (hPank4).
GN PANK4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

RNA SEQUENCE FROM N.A.
RA Isoqai T., Oca T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.,
RT "NEDO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
[2]
RNA SEQUENCE OF 6-773 FROM N.A.
RC TISSUE=Brain;
RA Wambutt R., Heubner D., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
[3]
RNA TISSUE SPECIFICITY.
RX MEDLINE:21372465; PubMed=11479594;
RA Zhou B., Westaway S.K., Levinson B., Johnson M.A., Gitschier J.,
RA Hayflick S.J.;
RT "A novel pantothenate kinase gene (PANK2) is defective in
RT Hallervorden-Spatz syndrome."
RL Nat. Genet. 28:345-349(2001).
CC -!- FUNCTION: Plays a role in the physiological regulation of the
CC intracellular CoA concentration (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + pantothenate = ADP + D-4'-
CC phosphopantothenate.
CC -!- ENZYME REGULATION: Regulated by feedback inhibition by CoA and its
CC thioesters (By similarity).
CC -!- PATHWAY: Coenzyme A (CoA) biosynthesis; first step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- TISSUE SPECIFICITY: Ubiquitously expressed with higher expression
CC in the muscle.
CC -!- SIMILARITY: BELONGS TO THE EUKARYOTIC PANTOTHENATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AK001644; BAA31805.1; -.
DR EMBL; AL442072; CAC09438.1; -.
DR Genes; HGNC:19366; PANK4.
DR MIM; 606162; -.
DR InterPro; IPR002791; DUF89.
DR InterPro; IPR004567; Pank_eukar.
DR Pfam; PF03630; Fumble; 1.
DR TIGRFAMs; TIGR00555; pank_eukar; 1.
KW Transferase; Kinase; ATP-binding; Coenzyme A biosynthesis;
KW Polymorphism.
FT VARIANT 684 684 Q -> R (IN dbSNP:2494620).
FT /FTID=VAR_015170.
FT CONFLICT 540 547 A -> V (IN REF. 2).
SQ SEQUENCE 773 AA; 85990 MW; D55027171A85E8B6 CRC64;
Query Match 30.0%; Score 6; DB 1; Length 773;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 YSWGEND 10
Db 377 YSWGEND 382
RESULT 9
ID PNK4 RAT STANDARD; PRT; 773 AA.
AC Q92358;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pantothenate kinase 4 (EC 2.7.1.33) (Pantothenic acid kinase 4)

DE (rPank4).
GN PANK4 OR FANG1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RA Chang Y.Y., Zuo J.J., Fang F.F.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Plays a role in the physiological regulation of the
CC intracellular CoA concentration (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + pantothenate = ADP + D-4'-
CC phosphopantothenate.
CC -!- ENZYME REGULATION: Regulated by feedback inhibition by CoA and its
CC thioesters (By similarity).
CC -!- PATHWAY: Coenzyme A (CoA) biosynthesis; first step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: BELONGS TO THE EUKARYOTIC PANTOTHENATE KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF399873; AAK94009.1; -;
CC InterPro; IPR004567; Pank_eukar.
CC Pfam; PF01937; DUF89; 1.
CC Pfam; PF03630; Fumble; 1.
CC TIGRFAMs; TIGR00555; pank_eukar; 1.
CC Trasnferase; Kinase; ATP-binding; Coenzyme A biosynthesis.
CC SEQUENCE 773 AA; 86242 MW; E56662F29616B40 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 773;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 YSWGEN 10
Db 377 YSWGEN 382

RESULT 10
ID MUDD CHLPN STANDARD; PRT; 809 AA.
AC Q9Z701; Q9Z0B5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE MurC/ddl bifunctional enzyme (includes: UDP-N-acetylmuramate--L-
DE alanine ligase (EC 6.3.2.8) (UDP-N-acetylmuramoyl-L-alanine
DE synthetase); D-alanine--D-alanine ligase (EC 6.3.2.4) (D-alanylalanine
DE synthetase) (D-Ala-D-Ala ligase)].
GN MURC/DDL OR CPN0905 OR CP0961.
OS Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."; Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;

RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.L.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39."; Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA."; Nucleic Acids Res. 28:2311-2314(2000).
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -!- FUNCTION: CELL WALL FORMATION.
CC -!- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanine.
CC phosphate + UDP-N-acetylmuramoyl-L-alanine = ADP + phosphate + D-
CC -!- CATALYTIC ACTIVITY: ATP + 2 D-alanine = ADP + phosphate + D-
CC alanyl-D-alanine.
CC -!- PATHWAY: Peptidoglycan biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: In the N-terminal section; belongs to the murCDEF
CC family.
CC -!- SIMILARITY: In the C-terminal section; belongs to the D-alanine--
CC D-alanine ligase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001670; AAD19043.1; -;
CC EMBL; AE002254; AAF38741.1; -;
CC EMBL; AP002548; BAA99113.1; -;
CC PIR; D72022; D72022.
CC PIR; G86603; G86603.
CC HSSP; P07862; LIOW.
CC PHCI-2DPAGE; Q9Z701; -;
CC TIGR; CP0961; -; fused; 1.
CC HAMAP; MF_00046; fused; 1.
CC InterPro; IPR005905; D_ala_D_ala.
CC InterPro; IPR000291; Dala_lig_Van.
CC InterPro; IPR000713; Mur_ligase.
CC InterPro; IPR004101; Mur_ligase_C.
CC InterPro; IPR005758; MurC.
CC Pfam; PF01820; Dala_Dala_ligase; 1.
CC Pfam; PF01225; Mur_ligase; 1.
CC Pfam; PF02875; Mur_ligase_C; 1.
CC TIGRFAMs; TIGR01205; D_ala_D_alatIGR; 1.
CC TIGRFAMs; TIGR01082; murC; 1.
CC PROSITE; PS00843; DALA_DALA_LIGASE_1; 1.
CC PROSITE; PS00844; DALA_DALA_LIGASE_2; FALSE NEG.
CC Peptidoglycan synthesis; Cell wall; Cell division; Ligase;
CC ATP-binding; Multifunctional enzyme; Complete proteome.
FT DOMAIN 1 450 UDP-N-ACETYLMURAMATE--ALANINE LIGASE.
FT NP_BIND 451 809 D-ALANINE--D-ALANINE LIGASE.
FT NP_BIND 111 117 ATP (POTENTIAL).
SQ SEQUENCE 809 AA; 89963 MW; 1A204C6B20E03B47 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 809;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 DTDVFEV 16
Db 646 DTDVFEV 651

RESULT 11
 CADI_XENLA
 ID_CADI_XENLA STANDARD; PRT; 872 AA.
 AC P30944; Q91709;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Epithelial-cadherin precursor (E-cadherin) (Uvomorulin) (XTCAD-1).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=83355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95324376; PubMed=7600966;
 RA Levine E., Lee C.H., Kintner C., Gumbiner B.M.;
 RT "Selective disruption of E-cadherin function in early Xenopus embryos
 by a dominant negative mutant";
 RL Development 120:901-909(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC Tissue-Tail bud;
 RX MEDLINE=94368839; PubMed=8086449;
 RA Tool O., Fujii G., Tashiro K., Shikawa K.;
 RT "Molecular cloning of cDNA for XTCAD-1, a novel Xenopus cadherin, and
 its expression in adult tissues and embryos of Xenopus laevis";
 RL Biochim. Biophys. Acta 1219:121-128(1994).
 RN [3]
 RP SEQUENCE OF 149-872 FROM N.A.
 RX MEDLINE=94363396; PubMed=8081882;
 RA Broders F., Girault J.M., Simonneau L., Thierry J.P.;
 RT "Sequence and distribution of Xenopus laevis E-cadherin transcripts";
 RL Cell Adhes. Commun. 1:265-277(1993).
 RN [4]
 RP SEQUENCE OF 149-169.
 RX MEDLINE=91347911; PubMed=1879345;
 RA Andres B., Mueller A.H.J., Kellermann J., Hausen P.;
 RT "Differential expression of two cadherins in Xenopus laevis";
 RL Development 111:829-844(1991).
 CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
 CC SORTING OF HETEROGENEOUS CELL TYPES. E-CADHERIN IS A LIGAND FOR
 CC INTEGRIN ALPHA-E/BETA-7.
 CC -!- SUBUNIT: Homodimer.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: NON-NEURAL EPITHELIAL TISSUES.
 CC -!- DEVELOPMENTAL STAGE: APPEARS IN THE EMBRYONIC ECTODERM DURING
 CC GASTRULATION WHEN EPIDERMAL DIFFERENTIATION COMMENCES AND IT
 CC DISAPPEARS FROM THE NEURAL PLATE AREA UPON NEURAL INDUCTION.
 CC -!- SIMILARITY: Contains 5 cadherin domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U04708; AAA93116.1; -;
 CC EMBL; L29057; AAA61489.1; -;
 CC EMBL; X75454; AAA53206.1; -;
 CC HSSP; P09803; 1SUH.
 CC InterPro; IPR002126; Cadherin.
 CC InterPro; IPR002233; Cadherin_C_term.
 CC Pfam; PF00028; cadherin; 5.
 CC Pfam; PF01049; Cadherin_C_term; 1.
 CC PRINTS; PR00205; CADHERIN.
 CC SMART; SM00112; CA; 4.

DR PROSITE; PS00232; CADHERIN 1; 3.
 DR PROSITE; PS0268; CADHERIN 2; 4.
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
 KW Signal.
 FT SIGNAL. 1 25 POTENTIAL.
 FT PROPEP 26 148
 FT CHAIN 149 872 EPITHELIAL-CADHERIN.
 FT DOMAIN 149 701 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 702 722 POTENTIAL.
 FT DOMAIN 723 872 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 244 254 CADHERIN 1.
 FT DOMAIN 358 368 CADHERIN 2.
 FT DOMAIN 577 587 CADHERIN 3.
 FT DOMAIN 710 721 POLY-LEU.
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 456 456 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 552 552 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 669 669 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 242 242 I -> V (IN REF. 3).
 FT CONFLICT 332 332 I -> T (IN REF. 3).
 FT CONFLICT 487 487 E -> R (IN REF. 3).
 FT CONFLICT 501 502 AT -> CS (IN REF. 3).
 FT CONFLICT 539 541 GNG -> EMA (IN REF. 3).
 FT CONFLICT 548 548 K -> R (IN REF. 3).
 FT CONFLICT 557 557 V -> G (IN REF. 3).
 FT CONFLICT 567 567 P -> L (IN REF. 3).
 FT CONFLICT 602 604 GFR -> EPQ (IN REF. 3).
 FT CONFLICT 638 640 GQS -> DK (IN REF. 2).
 FT CONFLICT 647 647 SILELRPK -> VYLSSDL (IN REF. 3).
 FT CONFLICT 660 660 T -> A (IN REF. 3).
 FT CONFLICT 770 770 A -> S (IN REF. 3).
 FT CONFLICT 842 842 P -> S (IN REF. 2 AND 3).
 FT CONFLICT 870 871 DE -> GED (IN REF. 3).
 SQ SEQUENCE 872 AA; 08637967EEFB4664 CRC64;
 Query Match 30.0%; Score 6; DB 1; Length 872;
 Best Local Similarity 100.0%; Pred.No. 44;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 15 FVLNNT 20
 DB 549 FVLNNT 554
 RESULT 12
 ANYG_CANAL STANDARD; PRT; 946 AA.
 ID ANYG_CANAL
 AC O74254;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glucoamylase I precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
 DE (1,4-alpha-D-glucan glucohydrolase).
 OS GAM1 OR GCAL.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=SC5314;
 RX MEDLINE=99451422; PubMed=10520161;
 RA Sturtevant J., Dixon F., Wadsworth E., Latge J.-P., Zhao X.-J.,
 RA Calderone R.;
 RT "Identification and cloning of GCAL, a gene that encodes a cell
 RT surface glucoamylase from Candida albicans";
 RL Med. Mycol. 37:357-366(1999).
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 CC glucose residues successively from non-reducing ends of the chains
 CC with release of beta-D-glucose.
 CC -!- SUBCELLULAR LOCATION: CELL WALL ASSOCIATED.
 CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF082188; AAC31968.1; -.
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Cell wall.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 946 GLUCOAMYLASE 1.
FT ACT_SITE 462 462 BY SIMILARITY.
FT DOMAIN 519 532 SER/THR-RICH.
FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 244 244 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 373 373 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 393 393 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 505 505 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 570 570 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 772 772 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 801 801 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 895 895 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 912 912 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 946 AA; 105804 MW; BD6B40C4EEF1F70 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 946;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 TDVFLV 17
DB 118 TDVFLV 123

RESULT 13
YJ9H YEAST
ID YJ9H YEAST STANDARD; PRT; 1648 AA.
AC P47171;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Hypothetical 191.7 kDa protein in H0M6-PMT4 intergenic region.
GN YJ9140C OR J2161.
OS Saccharomycetes cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Rose M., Koetter P., Entian K.D.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-1646 FROM N.A.
RA Scarcez T.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: SOME, TO S.POMBE SPBC31F10.14C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z49640; CAA89672.1; -.
DR EMBL; Z49641; CAA89674.1; -.
DR PIR; S57163; S57163.
DR TRANSFAC; T03439; -.
DR SGD; S0003901; HIR3.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0003714; F:transcription co-repressor activity; IMP.
DR GO; GO:0000083; P:G1/S-specific transcription in mitotic cell. .; IMP.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 391 411 POTENTIAL.
FT TRANSMEM 535 555 POTENTIAL.
FT TRANSMEM 563 583 POTENTIAL.
FT TRANSMEM 740 760 POTENTIAL.
FT TRANSMEM 803 823 POTENTIAL.
FT TRANSMEM 1461 1481 POTENTIAL.
FT TRANSMEM 1569 1589 POTENTIAL.
SQ SEQUENCE 1648 AA; 191679 MW; B1F31815797852D4 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 1648;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 DVFVLN 18
DB 1274 DVFVLN 1279

RESULT 14
AROQ AMYME
ID AROQ AMYME STANDARD; PRT; 20 AA.
AC P46380;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 3-dehydroquininate dehydratase (EC 4.2.1.10) (3-dehydroquinase)
DE (type II DHQase) (Fragment).
GN AROQ.
OS Amycolatopsis methanolica.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.
OX NCBI_TaxID=1814;
RN [1]
RP SEQUENCE.
RC STRAIN=NCIB 11946;
RX MEDLINE=93123995; PubMed=1479361;
RA Eversink G.J.W., Hessele G.I., Vrijbloed J.W., Coggins J.R.,
RA Dijkhuizen L.;
RT "Purification and characterization of a dual function
RT 3-dehydroquininate dehydratase from Amycolatopsis methanolica.";
RL J. Gen. Microbiol. 138:2449-2457(1992).
CC -1- FUNCTION: CATALYZE A TRANS-DEHYDRATION VIA AN ENOLATE
CC INTERMEDIATE. IS INVOLVED IN BOTH THE CATABOLISM OF QUINATE AND
CC THE BIOSYNTHESIS OF AROMATIC AMINO ACIDS. HAS A TEMPERATURE
CC OPTIMAL OF 76 DEGREES CELSIUS AND PH OPTIMAL OF 9.0. TYPE II
CC ENZYMES ARE THERMOSTABLE.
CC -1- CATALYTIC ACTIVITY: 3-dehydroquininate = 3-dehydroshikimate + H(2)O.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC third step.
CC -1- SUBUNIT: HOMODODECAMER.
CC -1- SIMILARITY: Belongs to the type-II 3-dehydroquinase family.
DR PIR; A47687; A47687.
DR HAMAP; MF 00169; -.
DR InterPro; IPR001874; DHQinase II.
DR Pfam; PF01220; DHQinase II; 1.
DR PROSITE; PS01029; DEHYDROQUINASE_II; PARTIAL.
KW Quinate metabolism; Aromatic amino acid biosynthesis; Lyase.
NON_TER 20 20
SQ SEQUENCE 20 AA; 2197 MW; C24AA183E5CFDFOA CRC64;

```

Query Match 25.0%; Score 5; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 VFVLN 18
DB 3 VFVLN 7

QY 12 TDVFFV 16
DB 24 TDVFFV 28

Search completed: November 21, 2003, 20:59:58
Job time : 5.85 secs

RESULT 15
CSP2_LACPL STANDARD; PRT; 66 AA.
AC P96349;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cold shock protein 2.
GN CSPL OR LP_0031.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LP80;
RX MEDLINE=97284511; PubMed=9139925;
RA Mayo B., Derzelle S., Fernandez M., Leonard C., Ferain T., Hols P.,
RA Suarez J.E., Delcour J.;
RT "Cloning and characterization of cspl and cspp, two cold-inducible
RT genes from Lactobacillus plantarum.";
RL J. Bacteriol. 179:3039-3042(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 8826 / WCFS1;
RX MEDLINE=22480296; PubMed=12566566;
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA Fieze M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
RA De Vos W.M., Siezen R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCFS1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- INDUCTION: In response to low temperature.
CC -!- SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y08940; CAA70139.1; -.
CC EMBL; AL935252; CAD62730.1; -.
CC HSSB; P15277; 1MJC.
CC InterPro; IPR002059; Cold_shock.
CC Pfam; PF00313; CSD; 1.
CC PRINTS; PR00050; COLDSHOCK.
CC ProDom; PD000621; Cold shock; 1.
CC SMART; SM00357; CSP; 1.
CC PROSITE; PS00352; COLD_SHOCK; 1.
KW Transcription regulation; DNA-binding; Activator; Multigene family;
FT Complete proteome.
FT DOMAIN 4 63 CSD.
SQ SEQUENCE 66 AA; 7300 MW; 933C202684587579 CRC64;

Query Match 25.0%; Score 5; DB 1; Length 66;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: November 21, 2003, 20:45:57 ; Search time 24.4 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-82
Perfect score: 20
Sequence: 1 GAPTYSWGENDTVFVLNLT 20
Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 830525 seqs, 258052604 residues
Word size : 0
Total number of hits satisfying chosen parameters: 830525
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	2436	12 Q81756	Q81756 hepatitis c
2	20	100.0	3011	12 Q91F55	Q91F55 hepatitis c
3	14	70.0	137	12 Q914Q6	Q914Q6 hepatitis c
4	14	70.0	137	12 Q914Q5	Q914Q5 hepatitis c
5	14	70.0	137	12 Q914Q8	Q914Q8 hepatitis c
6	14	70.0	137	12 Q914Q4	Q914Q4 hepatitis c
7	14	70.0	137	12 Q914Q0	Q914Q0 hepatitis c
8	14	70.0	137	12 Q914Q3	Q914Q3 hepatitis c
9	14	70.0	137	12 Q914Q2	Q914Q2 hepatitis c
10	14	70.0	137	12 Q914Q7	Q914Q7 hepatitis c
11	14	70.0	746	12 Q8JPM2	Q8JPM2 hepatitis c
12	12	60.0	1275	12 Q06642	Q06642 hepatitis c
13	12	60.0	3015	12 Q9WPH5	Q9WPH5 hepatitis c
14	11	55.0	137	12 Q914W1	Q914W1 hepatitis c
15	11	55.0	137	12 Q914W4	Q914W4 hepatitis c
16	11	55.0	137	12 Q914W9	Q914W9 hepatitis c

17	11	55.0	137	12 Q914W2	Q914W2 hepatitis c
18	11	55.0	137	12 Q914W8	Q914W8 hepatitis c
19	11	55.0	137	12 Q914W5	Q914W5 hepatitis c
20	11	55.0	137	12 Q914W0	Q914W0 hepatitis c
21	11	55.0	137	12 Q914X0	Q914X0 hepatitis c
22	11	55.0	137	12 Q914V9	Q914V9 hepatitis c
23	11	55.0	137	12 Q914V8	Q914V8 hepatitis c
24	11	55.0	137	12 Q914W6	Q914W6 hepatitis c
25	11	55.0	137	12 Q914W7	Q914W7 hepatitis c
26	11	55.0	137	12 Q914V7	Q914V7 hepatitis c
27	11	55.0	137	12 Q914W3	Q914W3 hepatitis c
28	11	55.0	206	12 Q81571	Q81571 hepatitis c
29	11	55.0	403	12 Q9PX22	Q9PX22 hepatitis c
30	11	55.0	640	12 Q68966	Q68966 hepatitis c
31	11	55.0	778	12 Q04184	Q04184 hepatitis c
32	11	55.0	3011	12 Q36579	Q36579 hepatitis c
33	11	55.0	3011	12 Q36610	Q36610 hepatitis c
34	11	55.0	3011	12 Q03463	Q03463 hepatitis c
35	11	55.0	3011	12 Q9EL58	Q9EL58 hepatitis c
36	11	55.0	3011	12 Q36609	Q36609 hepatitis c
37	11	55.0	3011	12 Q9DIT6	Q9DIT6 hepatitis c
38	11	55.0	3011	12 Q36608	Q36608 hepatitis c
39	10	50.0	187	12 Q9PXG5	Q9PXG5 hepatitis c
40	10	50.0	279	12 Q68457	Q68457 hepatitis c
41	10	50.0	279	12 Q68458	Q68458 hepatitis c
42	10	50.0	279	12 Q68456	Q68456 hepatitis c
43	10	50.0	279	12 Q81324	Q81324 hepatitis c
44	10	50.0	350	12 Q81324	Q81324 hepatitis c
45	10	50.0	364	12 Q36980	Q36980 hepatitis c

ALIGNMENTS

RESULT 1

Q81756 PRELIMINARY; PRT; 2436 AA.
ID Q81756
AC Q81756;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Choo Q.-L., Richman K., Han J.;
RT "The nucleotide sequence of the Hepatitis C viral genome."
RL Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.
DR EMBL; M32084; AAA45677.1; -.
DR HSSP; P27958; IALV.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR007095; RNA pol Ds PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; helicase C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR ProDom; PD186062; HCV NS1; 1.

```
DR SMART: SM00487; DEXDC; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
FT NON_TER 1 2436
FT NON_TER 2436 AA; 264734 MW; D7B9872900BE3125 CRC64;
SQ SEQUENCE 2436 AA; 264734 MW; D7B9872900BE3125 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 2436;
Best Local Similarity 100.0%; Pred. No. 4.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDTDVFLNNT 20
Db 73 GAPTYSWGENDTDVFLNNT 92

RESULT 2
Q91FES PRELIMINARY; PRT; 3011 AA.
AC Q91FES
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype sequence.";
RL J. Gen. Virol. 82:1291-1297(2001).
CC -1. SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND MRNA (BY SIMILARITY).
CC EMBL; AF271632; AAF81759.1; -.
DR HSP; P27958; 1A1V.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV env.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RDRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01339; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00398; Viral_RDRP; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
```

```
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327124 MW; 2489CE74AC864E58 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAPTYSWGENDTDVFLNNT 20
Db 523 GAPTYSWGENDTDVFLNNT 542

RESULT 3
Q914Q6 PRELIMINARY; PRT; 137 AA.
AC Q914Q6
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=wt;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431882; AAL30731.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1 137
FT NON_TER 137 137
SQ SEQUENCE 137 AA; 15042 MW; 08E5232119C91A09 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 WGENDTDVFLNNT 20
Db 51 WGENDTDVFLNNT 64

RESULT 4
Q914Q5 PRELIMINARY; PRT; 137 AA.
AC Q914Q5
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=wt;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
```

```

DR EMBL; AF431883; AAL30732.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 15070 MW; 08F2132F29D01A09 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 WGENDTDFVLNNT 20
DB 51 WGENDTDFVLNNT 64

RESULT 5
ID Q914Q8 PRELIMINARY; PRT; 137 AA.
AC Q914Q8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=wt;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431880; AAL30729.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 15070 MW; 08F2132F29D01A09 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 WGENDTDFVLNNT 20
DB 51 WGENDTDFVLNNT 64

RESULT 6
ID Q914Q4 PRELIMINARY; PRT; 137 AA.
AC Q914Q4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=wt;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;

```

```

RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431884; AAL30733.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 15070 MW; 08F2132F29D01A09 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 WGENDTDFVLNNT 20
DB 51 WGENDTDFVLNNT 64

RESULT 7
ID Q914Q0 PRELIMINARY; PRT; 137 AA.
AC Q914Q0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=wt;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431888; AAL30737.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 15078 MW; FF7E4F2219D33F35 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 WGENDTDFVLNNT 20
DB 51 WGENDTDFVLNNT 64

RESULT 8
ID Q914Q3 PRELIMINARY; PRT; 137 AA.
AC Q914Q3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]

```

```
RP SEQUENCE FROM N.A.
RC STRAIN=wt;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431885; AAL30734.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV_NS1; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 15074 MW; 18B885A375CC1A09 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 WGENDTDFVLNNT 20
Db 51 WGENDTDFVLNNT 64

RESULT 9
Q914Q2 ID Q914Q2 PRELIMINARY; PRT; 137 AA.
AC Q914Q2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=wt;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431886; AAL30735.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV_NS1; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 15070 MW; 03A3EF7374DC0B08 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 WGENDTDFVLNNT 20
Db 51 WGENDTDFVLNNT 64

RESULT 10
Q914Q7 ID Q914Q7 PRELIMINARY; PRT; 137 AA.
AC Q914Q7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
```

```
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=wt;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431881; AAL30730.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV_NS1; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 15070 MW; 08F2132F29D01A09 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 WGENDTDFVLNNT 20
Db 51 WGENDTDFVLNNT 64

RESULT 11
Q8JPM2 ID Q8JPM2 PRELIMINARY; PRT; 746 AA.
AC Q8JPM2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21904749; PubMed=11907246;
RA Blanchard E., Brand D., Trassard S., Goudeau A., Roingeard P.;
RT "Hepatitis C virus-like particle morphogenesis.";
RL J. Virol. 76:4073-4079(2002).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND RNA (BY SIMILARITY).
CC EMBL; AF529293; AM94419.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV_NS2.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 746
SQ SEQUENCE 746 AA; 82116 MW; 2DA94CC6845BA92C CRC64;

Query Match 70.0%; Score 14; DB 12; Length 746;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 WGENDTDFVLNNT 20
Db 51 WGENDTDFVLNNT 64
```



```
Db 529 WGENDTDFVFLNNT 542
RESULT 12
Q06642 PRELIMINARY; PRT; 1275 AA.
ID Q06642
AC Q06642;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE HCV polyprotein (Genome polyprotein) (Fragment).
OS Hepatitis C virus type 2.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=40271;
RN [1]
RP SEQUENCE FROM N.A.
RA Hayashi N., Higashi H., Kaminaka K., Sugimoto H., Esumi M.,
RA Komatsu K., Hayashi K., Sugitani M., Suzuki K., Okano T., Nozaki C.,
RA Mizuno K., Shikata T.;
RT "Molecular cloning and heterogeneity of the human hepatitis C virus
RT (HCV) genome.";
RL J. Hepatol. 17:0-0(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93314964; PubMed=7686870;
RA Nishihara T., Nozaki C., Nakatake H., Hoshiko K., Esumi M.,
RA Hayashi N., Hino K., Hanada P., Mizuno K., Shikata T.;
RT "Immunoreactive core peptides of hepatitis C virus produced in
RT Escherichia coli and in vitro DNA amplification-restricted
RT transcription-translation system.";
RL J. Virol. Methods 59:91-98(1996).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=96386004; PubMed=8793834;
RA Esumi M., Hayashi N., Takahashi H., Shikata T., Moriyama M.,
RA Arakawa Y., Eco T., Nishihara T., Nozaki C., Mizuno K.;
RT "Reactive capture hepatitis C virus.";
RL Virology 251:158-164(1998).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; D13406; BAA02670.1; -.
DR HSP; P26683; INS3.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR002518; HCV NS2.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1275
SQ SEQUENCE 1275 AA; 138073 MW; F0D4F26618994B46 CRC64;

Query Match 60.0%; Score 12; DB 12; Length 1275;
Best Local Similarity 100.0%; Pred. No. 3.2e-05;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 PTYSWGENDTDV 14
Db 526 PTYSWGENDTDV 537
|||||
|||||

RESULT 13
Q9WPH5 PRELIMINARY; PRT; 3015 AA.
ID Q9WPH5
AC Q9WPH5;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-N.
RX MEDLINE=99315771; PubMed=10385673;
RA Beard M.R., Abell G., Honda M., Carroll A., Gartland M., Clarke B.,
RA Suzuki K., Lanford R., Sangar D.V., Lemon S.M.;
RT "An infectious molecular clone of a Japanese genotype 1b hepatitis C
RT virus.";
RL Hepatology 30:316-324(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-N;
RX MEDLINE=99315771; PubMed=10385673;
RA Beard M.R., Abell G., Honda M., Carroll A., Gartland M., Clarke B.,
RA Suzuki K., Lanford R., Sangar D.V., Lemon S.M.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF139594; AAD44718.2; -.
SQ SEQUENCE 3015 AA; 863623 MW; A86AE71196578EE3 CRC64;

Query Match 60.0%; Score 12; DB 12; Length 3015;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 PTYSWGENDTDV 14
Db 526 PTYSWGENDTDV 537
|||||
|||||

RESULT 14
Q914W1 PRELIMINARY; PRT; 137 AA.
ID Q914W1
AC Q914W1;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=sg;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
RT and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431826; AAL30676.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 137
SQ SEQUENCE 137 AA; 138073 MW; F0D4F26618994B46 CRC64;
```

SQ SEQUENCE 137 AA; 14996 MW; 46F1541F17CFF13F CRC64;

Query Match 55.0%; Score 11; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 6e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 NDTDVFLNNT 20
|||||
Db 54 NDTDVFLNNT 64

RESULT 15

Q914W4 PRELIMINARY; PRT; 137 AA.

AC Q914W4;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=sg;
RA Lyra A.C.; Fan X.; Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431823; AAL30673.1; -;
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137 137
SQ SEQUENCE 137 AA; 14989 MW; 0C203D19D57247E4 CRC64;

Query Match 55.0%; Score 11; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 6e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 NDTDVFLNNT 20
|||||
Db 54 NDTDVFLNNT 64

Search completed: November 21, 2003, 21:08:19
Job time : 25.45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:22 ; Search time 32.1 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-83

Perfect score: 20

Sequence: 1 GNWFGCTWNSGTGTFKVCQA 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	16 AAR84512	Hepatitis C virus
2	20	100.0	20	17 AAR91006	HCV E2 peptide E2-
3	20	100.0	20	23 AAO18712	Hepatitis C virus
4	20	100.0	26	13 AAR20752	Peptide 2 based on
5	20	100.0	30	23 AAU84634	HCV HepC1a segment
6	20	100.0	96	23 ABB77258	HCV bait polypepti
7	20	100.0	176	23 ABB77257	HCV bait polypepti
8	20	100.0	179	21 AAB18529	Protein encoded by
9	20	100.0	192	19 AAW67009	HCV nucleocapsid C

10	20	100.0	250	23 AAU79220	Hepatitis C Virus
11	20	100.0	254	22 AAB68043	Amino acid sequenc
12	20	100.0	271	22 AAE00445	HCV E2-634 HVR1 co
13	20	100.0	271	22 AAE00446	HCV E2-634 HVR1-mu
14	20	100.0	278	14 AAR33997	Th E2/NS1 protein.
15	20	100.0	305	18 AAW00929	Recombinant HCV E2
16	20	100.0	333	14 AAR40118	HGH-HCV-E2 fusion
17	20	100.0	337	16 AAR79217	pHCV351-encoded AP
18	20	100.0	350	23 AAE19891	Hepatitis C virus
19	20	100.0	363	22 AAB68042	Amino acid sequenc
20	20	100.0	363	24 AAE32880	Hepatitis C virus
21	20	100.0	363	24 ABE55567	Hepatitis C virus
22	20	100.0	367	14 AAR40115	APP-HCV-E2 fusion
23	20	100.0	367	16 AAR79218	pHCV167-encoded pr
24	20	100.0	377	16 AAR79226	pHCV422-encoded AP
25	20	100.0	397	16 AAR79220	pHCV419-encoded AP
26	20	100.0	399	14 AAR40117	HGH-HCV-E2 fusion
27	20	100.0	402	14 AAR34438	Sequence of glycop
28	20	100.0	402	14 AAR34439	Sequence of glycop
29	20	100.0	402	14 AAR34440	Sequence of glycop
30	20	100.0	409	14 AAR33995	H77 E2/NS1 protein
31	20	100.0	409	14 AAR33996	H90 E2/NS1 protein
32	20	100.0	410	16 AAR79227	pHCV423-encoded AP
33	20	100.0	417	16 AAR79228	pHCV424-encoded AP
34	20	100.0	434	16 AAR79219	pHCV418-encoded AP
35	20	100.0	441	16 AAR79230	pHCV429-encoded AP
36	20	100.0	447	16 AAR79229	pHCV425-encoded AP
37	20	100.0	453	16 AAR79225	pHCV421-encoded AP
38	20	100.0	454	10 AAR90183	Sequence of hepati
39	20	100.0	454	10 AAR92049	Sequence encoded b
40	20	100.0	454	21 AAB18526	Protein encoded by
41	20	100.0	463	14 AAR33588	HCV CKS-NS1S1 fusi
42	20	100.0	463	14 AAR33187	Sequence of subtra
43	20	100.0	463	22 AAB69007	HCV recombinant an
44	20	100.0	480	14 AAR33992	HCV-1 E2/NS1 prote
45	20	100.0	490	16 AAR79224	pHCV420-encoded AP

ALIGNMENTS

RESULT 1
AAR84512
ID AAR84512 standard; peptide; 20 AA.
AC AAR84512;
XX
XX
DT 06-JAN-1997 (first entry)
XX
DE Hepatitis C virus peptide NS1-17 (residues 547-566).
XX
XX Hepatitis C virus; HCV; immunogen; non-structural region; NS1;
KW immunodominant; T cell epitope; vaccine.
XX
OS Hepatitis C virus.
XX
PN WO9512677-A2.
XX
PD 11-MAY-1995.
XX
PF 28-OCT-1994; 94WO-EP03555.
XX
PR 04-NOV-1993; 93EP-0402718.
XX
PA (INNO-) INNOGENETICS NV.
XX
PI Deleys R, Leroux-Roels G, Maertens G;
XX
DR WPI; 1995-193822/25.
XX
PT Hepatitis C Virus immunogenic polypeptide contg. a T-cell
stimulating epitope - from core, E1, E2 and NS3 regions, useful in
production of vaccines, therapeutic agents, etc.

XX Example 4; Page 51; 105pp; English.

PS A series of overlapping peptides (including the present sequence) was

CC synthesised based on sequences in the core, E1 and E2/NS1 regions of

CC hepatitis C virus. The peptides were used as antigens in lympho-

CC proliferative assays to identify the main T-cell epitopes.

XX Sequence 20 AA;

SQ Query Match 100.0%; Score 20; DB 16; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.8e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20

Db 1 GNWFGCTWMNSTGFTKVCGA 20

RESULT 2

AAR91006

ID AAR91006 standard; peptide; 20 AA.

XX AC AAR91006;

XX DT 25-SEP-1996 (first entry)

XX HCV E2 peptide E2-17B for competition studies.

DE HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;

KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.

KW Synthetic.

OS WO9604385-A2.

XX PN 15-FEB-1996.

XX PD 31-JUL-1995; 95WO-EP03031.

XX PF 29-JUL-1994; 94EP-0870132.

XX PR (INNO-) INNOGENETICS NV.

XX PA Bosman F, Buyse M, De Martynoff G, Maertens G;

XX PI WPI; 1996-129401/13.

XX DR Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope

XX PT proteins - in presence of disulphide bond cleavage agent, to

XX PT produce proteins suitable for direct use in vaccines or diagnostic

XX PT assays of HCV

XX PS Claim 29; Page 67; 146pp; English.

XX CC AAR90965-R90980 and AAR90995-R91015 represent synthetic hepatitis C

XX CC virus (HCV) E1 and E2 peptides used in competition studies. This

XX CC sequence represents a synthetic E2 peptide, and corresponds to residues

XX CC 547-566 of the E2 protein sequence. These sequences are useful for in

XX CC vitro monitoring of HCV disease, or prognosis of the response to

XX CC interferon treatment of patients suffering from HCV infection. These

XX CC sequences compete with the proteins produced by AAT12704-T12709 and

XX CC AAT12961-T12974, which are included in vectors for the production of

XX CC recombinant E1, E2, and E1/E2. The recombinant proteins can then be

XX CC isolated and purified by carrying out a disulphide bond cleavage, or a

XX CC reduction step with a disulphide bond cleavage agent, after lysis of

XX CC recombinant host cells. The constructs containing the purified HCV

XX CC envelope proteins can be used for vaccinating humans against HCV, for in

XX CC vitro detection of HCV antibodies in a sample, and in a serotyping assay

XX CC for detecting one or more serological types of HCV present in a

XX CC biological sample. The constructs can also be immobilised on a solid

XX CC substrate and incorporated into a reversed phase hybridisation assay for

XX CC determining the presence or the genotype of HCV. The new purification

CC method preserves the conformation of the recombinantly expressed E1, E2

CC and E1/E2, and eliminates contaminating proteins. Antigens isolated

CC using this method are more reactive with human sera than those isolated

CC by known techniques.

XX Sequence 20 AA;

SQ Query Match 100.0%; Score 20; DB 17; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.8e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20

Db 1 GNWFGCTWMNSTGFTKVCGA 20

RESULT 3

AAO18712

ID AAO18712 standard; Peptide; 20 AA.

XX AC AAO18712;

XX DT 24-OCT-2002 (first entry)

XX Hepatitis C virus E2 protein derived peptide E2-17B.

DE Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide;

KW immunostimulant; vaccine.

XX OS Hepatitis C virus.

XX PN WO200255548-A2.

XX PD 18-JUL-2002.

XX PF 11-JAN-2002; 2002WO-EP00219.

XX PR 11-JAN-2001; 2001US-260699P.

XX PR 30-AUG-2001; 2001US-315768P.

XX PA (INNO-) INNOGENETICS NV.

XX PI Maertens G, Bosman F, Buyse M;

XX DR WPI; 2002-599657/64.

XX PT New therapeutic vaccine compositions comprising at least one purified

XX PT recombinant hepatitis C virus (HCV) single or specific oligomeric

XX PT recombinant envelope protein E1 or E2, useful for immunizing humans

XX PT from HCV infection -

XX PS Claim 4; Page 229; 243pp; English.

XX CC The present invention relates to new therapeutic vaccine compositions for

XX CC inducing hepatitis C virus (HCV)-specific antibodies, comprising a

XX CC composition containing at least one purified recombinant HCV single or

XX CC specific oligomeric recombinant envelope proteins selected from an E1 and

XX CC an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are

XX CC useful for inducing HCV-specific antibodies or for immunising humans

XX CC against HCV. The recombinant HCV E1 and/or E2 proteins are useful as

XX CC vaccines or therapeutics, in HCV screening and confirmatory antibody

XX CC tests, for raising antibodies, in the preparation of medicament, and for

XX CC in vitro monitoring of HCV disease or prognosing the response to

XX CC treatment of patients suffering from HCV infection. The present sequence

XX CC is a peptide derived from the proteins of the invention.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 23; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.8e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20

DB 1 GNWFGCTWMNSTGFTKVCGA 20
|||||

RESULT 4
AAR20752
ID AAR20752 standard; Protein; 26 AA.
XX AC AAR20752;
XX 25-MAR-2003 (updated)
DT 05-MAY-1992 (first entry)
XX Peptide 2 based on immunoreactive region of Hepatitis C virus.
DE XX Non-A, non-B hepatitis virus; non-structural protein; vaccine.
XX Synthetic.
XX EP468527-A.
XX 29-JAN-1992.
XX 26-JUL-1991; 91EP-0112620.
XX 24-JUN-1991; 91US-0719819.
PR 26-JUL-1990; 90US-0558799.
PR 07-FEB-1991; 91US-0651735.
PR 11-MAR-1991; 91US-0667275.
XX (UNBI-) UNITED BIOMEDICAL INC.
PA Chang YW, Hossein B;
PI WPI; 1992-034279/05.
XX New synthetic peptide specific for HCV antibodies - for detection
PT of HCV or NANBHV e.g. by enzyme-linked immunosorbent assay and is
PT immunogen for preparation of vaccines
XX Claim 1; Page 88; 98pp; English.
XX This peptide is one of 19 specifically claimed antigens based on the
CC immunoreactive regions of the envelope protein and non-structural
CC proteins NS-1, NS-2, NS-3 and NS-5 for the Hepatitis C virus. The
CC invention also concerns analogues, segments, mixtures, conjugates
CC and polymers of these peptides. The C-terminal amino acid may be
CC amidated. See AAR20751-R20782.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX Sequence 26 AA;

Query Match 100.0%; Score 20; DB 13; Length 26;
Best Local Similarity 100.0%; Pred. No. 2.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||||
DB 4 GNWFGCTWMNSTGFTKVCGA 23

RESULT 5
AAU84634
ID AAU84634 standard; Peptide; 30 AA.
XX AC AAU84634;
XX 08-MAY-2002 (first entry)
XX HCV HepC1a segment 37.
XX Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
KW viral infection; human immunodeficiency virus; melanoma;

KW bacterial infection; Salmonella; Legionella; parasitic infection;
XX Trypanosoma; Toxoplasma; Giardia.
OS Hepatitis C virus.
XX WO200190197-A1.
PN 29-NOV-2001.
XX 25-MAY-2001; 2001WO-AU00622.
XX 26-MAY-2000; 2000AU-0007761.
PR (AUSU) UNIV AUSTRALIAN NAT.
XX Thomson SA, Ramshaw IA;
XX WPI; 2002-147575/19.
DR N-PSDB; ABK36472.

New synthetic polypeptides having several different segments of at least one parent polypeptide linked together differently compared to the linkage in the parent polypeptide, for inducing immune response against a pathogen or cancer -

Example 2; Fig 26; 364pp; English.

The invention relates to a new synthetic polypeptide (I) comprising several different segments of at least one parent polypeptide linked together in a different relationship relative to their linkage in the parent polypeptide to impede, abrogate or otherwise alter at least one function associated with the parent polypeptide and for inducing an immune response against a pathogen or cancer. Also included are a synthetic polynucleotide encoding and a computer system for designing the synthetic polypeptides. The synthetic polypeptides and polynucleotides are referred to as a Savine. The synthetic polypeptide is useful for modulating immune responses preferably directed against a pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix, colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone liver, oesophagus, brain, testicle, uterus), as potentiating agents. Compositions comprising the polypeptide may be used in the treatment or prophylaxis against viral (such as infections caused by HIV (human immunodeficiency virus), hepatitis, influenza, Japanese encephalitis virus, Epstein-Barr virus and respiratory syncytial virus), bacterial (e.g., infections caused by Neisseria, Meningococcal, Haemophilus, Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic (e.g., infections caused by Plasmodium, Schistosoma, Leishmania, Trypanosoma, Toxoplasma and Giardia) infections. The present sequence is a peptide derived from a parent protein used to construct a savine of the invention.

SQ Sequence 30 AA;

Query Match 100.0%; Score 20; DB 23; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||||
DB 9 GNWFGCTWMNSTGFTKVCGA 28

RESULT 6
ABB77258
ID ABB77258 standard; Protein; 96 AA.
XX ABB77258;
XX 28-JUN-2002 (first entry)
XX HCV bait polypeptide 6.
XX SID; selected interacting domain; HCV; hepatitis C virus;

KW liver disease; liver cancer; virucide; hepatotropic; antiinflammatory;
 KW antibacterial.

OS Hepatitis C virus strain H77.

XX EP1178116-A1.

XX PN 06-FEB-2002.

PD 03-AUG-2000; 2000EP-0402225.

XX 03-AUG-2000; 2000EP-0402225.

XX (HYBR-) HYBRIGENICS SA.

XX Legrain P, Whiteside S, Wojcik J;

XX WPI; 2002-208115/27.

XX N-PSDB; ABL55590.

XX New selected interacting domain polypeptides and polynucleotides,
 PT useful for treating or preventing infections or pathologies caused by
 PT hepatitis C virus (HCV) or those linked to HCV infection -

XX Claim 26; SEQ ID 82; 61pp + sequence listing; English.

XX The invention relates to nucleic acids encoding polypeptides which are
 CC termed SID polypeptides (selected interacting domain). These polypeptides
 CC are the final products of a double selection method involving a first
 CC step of selection of Hepatitis C virus (HCV)-derived polynucleotides
 CC through a two-hybrid system, and a second selection step involving an
 CC alignment between the different polynucleotides selected at the first
 CC step. The activity of polypeptides of the invention may be described as,
 CC virucide, hepatotropic, antiinflammatory and antibacterial. The
 CC polypeptide, polynucleotide and compositions comprising them are useful
 CC for treating or preventing viral or a bacterial infection, specifically
 CC infections or pathologies caused by HCV, or those pathologies linked to
 CC HCV infection. These may include liver disease and liver cancer. The
 CC current sequence represents a HCV bait polypeptide.

CC Note: The sequence data for this patent is not represented in the
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.

XX Sequence 96 AA;

Query Match 100.0%; Score 20; DB 23; Length 96;
 Best Local Similarity 100.0%; Pred. No. 6.5e-14;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVGCA 20

Db 61 GNWFGCTWMNSTGFTKVGCA 80

RESULT 7

ABB77257

ID ABB77257 standard; Protein; 176 AA.

XX ABB77257;

XX 28-JUN-2002 (first entry)

XX HCV bait polypeptide 5.

XX SID; selected interacting domain; HCV; hepatitis C virus;
 KW liver disease; liver cancer; virucide; hepatotropic; antiinflammatory;
 KW antibacterial.

XX Hepatitis C virus strain H77.

XX EP1178116-A1.

XX 06-FEB-2002.

XX 03-AUG-2000; 2000EP-0402225.

XX 03-AUG-2000; 2000EP-0402225.

XX (HYBR-) HYBRIGENICS SA.

XX Legrain P, Whiteside S, Wojcik J;

XX WPI; 2002-208115/27.

XX N-PSDB; ABL55589.

XX New selected interacting domain polypeptides and polynucleotides,
 PT useful for treating or preventing infections or pathologies caused by
 PT hepatitis C virus (HCV) or those linked to HCV infection -

XX Claim 26; SEQ ID 81; 61pp + sequence listing; English.

XX The invention relates to nucleic acids encoding polypeptides which are
 CC termed SID polypeptides (selected interacting domain). These polypeptides
 CC are the final products of a double selection method involving a first
 CC step of selection of Hepatitis C virus (HCV)-derived polynucleotides
 CC through a two-hybrid system, and a second selection step involving an
 CC alignment between the different polynucleotides selected at the first
 CC step. The activity of polypeptides of the invention may be described as,
 CC virucide, hepatotropic, antiinflammatory and antibacterial. The
 CC polypeptide, polynucleotide and compositions comprising them are useful
 CC for treating or preventing viral or a bacterial infection, specifically
 CC infections or pathologies caused by HCV, or those pathologies linked to
 CC HCV infection. These may include liver disease and liver cancer. The
 CC current sequence represents a HCV bait polypeptide.

CC Note: The sequence data for this patent is not represented in the
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.

XX Sequence 176 AA;

Query Match 100.0%; Score 20; DB 23; Length 176;
 Best Local Similarity 100.0%; Pred. No. 1.1e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVGCA 20

Db 79 GNWFGCTWMNSTGFTKVGCA 98

RESULT 8

AAB18529

ID AAB18529 standard; Protein; 179 AA.

XX AAB18529;

XX 15-JAN-2001 (first entry)

XX Protein encoded by a novel hepatitis C virus cDNA clone 131.

XX Hepatitis C virus; HCV; antisense polynucleotide; polypeptide;
 KW viral infectivity; viral replication.

XX Hepatitis C virus.

XX EP1034785-A2.

XX 13-SEP-2000.

XX 16-MAR-1990; 2000EP-0109602.

XX 17-MAR-1989; 89US-0325338.

XX 20-APR-1989; 89US-0341334.

XX 18-MAY-1989; 89US-0355002.

XX 16-MAR-1990; 90EP-0302866.

XX (CHIR) CHIRON CORP.

XX Houghton M, Choo Q, Kuo G;
 XX WPI; 2000-566891/53.
 DR N-PSDB; AAA75285.
 XX Novel composition comprising a hepatitis C virus antisense
 PT polynucleotide which is complementary to or corresponds to a sense
 PT strand of the virus genome, and selectively hybridises to it -
 XX
 PS Example; Fig 5; 75pp; English.
 XX The specification describes a pharmaceutical composition which
 CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
 CC HCV is characterized by a positive stranded RNA genome which has
 CC 40% homology at the polypeptide level to a HCV polypeptide. The
 CC antisense polynucleotide binds to cellular polynucleotides which
 CC enhance and/or are required for viral infectivity, replicative
 CC ability or chronicity. The antisense polynucleotides may also be
 CC designed to bind with high specificity, to be of increased stability,
 CC to be stable and to have low toxicity. The composition also comprises
 CC an agent which causes viral RNA to be inactive. The composition
 CC is used for preventing HCV replication in a system. The present
 CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
 CC course of the invention.
 XX
 XX Sequence 179 AA;
 SQ
 Query Match 100.0%; Score 20; DB 21; Length 179;
 Best Local Similarity 100.0%; Pred. No. 1.1e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GNWFGCTWMNSTGFTKVCGA 20
 DB 37 GNWFGCTWMNSTGFTKVCGA 56
 RESULT 9
 AA67009
 ID AA67009 standard; protein; 192 AA.
 XX
 AC AA67009;
 XX
 DT 02-MAR-1999 (first entry)
 XX
 DE HCV nucleocapsid core protein.
 XX
 KW Hepatitis C virus; HCV; nucleocapsid; core protein; analogue; antibody;
 KW non-structural protein; thioamide bond; peptide bond.
 XX
 OS Hepatitis C virus.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 79
 FT /note= "given in specification as Try"
 XX
 PN JP10226698-A.
 XX
 PD 25-AUG-1998.
 XX
 PF 19-FEB-1997; 97JP-0034702.
 XX
 PR 19-FEB-1997; 97JP-0034702.
 XX
 PA (KYOW) KYOWA MEDEX KK.
 XX
 DR WPI; 1998-515103/44.
 XX
 XX Determination of antibody in sample - uses peptide analog absorbed
 PT or chemically bound on carrier as antigen
 XX
 XX Disclosure; Page 4; 13pp; Japanese.
 XX

CC This sequence represents the Hepatitis C virus (HCV) nucleocapsid core
 CC protein. The invention relates to peptide analogues derived from HCV
 CC proteins, e.g. AA67417-W67426, which can be used for the determination
 CC of anti-HCV antibodies in a sample. Preferably the peptide analogues
 CC contain one or more thioamide peptide bonds where at least one oxygen
 CC atom of the peptide bond is replaced by sulphur atom. The peptide
 CC analogues can be adsorbed or chemically bound to a carrier.
 XX
 SQ Sequence 192 AA;
 Query Match 100.0%; Score 20; DB 19; Length 192;
 Best Local Similarity 100.0%; Pred. No. 1.2e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GNWFGCTWMNSTGFTKVCGA 20
 DB 97 GNWFGCTWMNSTGFTKVCGA 116
 RESULT 10
 AAU79220
 ID AAU79220 standard; Protein; 250 AA.
 XX
 AC AAU79220;
 XX
 DT 15-JUL-2002 (first entry)
 XX
 DE Hepatitis C Virus (HCV) delta-delta E2 genotype 1a protein.
 XX
 KW Hepatitis C Virus E2; HCV E2; virucide; hepatotropic; IL-2;
 KW antiinflammatory; HCV infection; interleukin-2; gamma-interferon;
 KW granulocyte macrophage-colony stimulating factor; GM-CSF;
 KW delta-delta E2 genotype 1a.
 XX
 OS Hepatitis C Virus.
 XX
 PN WO200222155-A1.
 XX
 PD 21-MAR-2002.
 XX
 PF 13-SEP-2001; 2001WO-US28767.
 XX
 PR 13-SEP-2000; 2000US-230927P.
 XX
 PA (HAWA-) HAWAII BIOTECHNOLOGY GROUP INC.
 XX
 PI Nakano ET, Clements DE, Humphreys T;
 XX
 DR WPI; 2002-383102/41.
 DR N-PSDB; ABK49386.
 XX
 PT New immunogenic polypeptide comprising hepatitis C virus E2 polypeptide
 PT useful for treating hepatitis C virus infection and for providing
 PT immune protection against virus infection -
 XX
 PS Claim 4; Fig 6; 84pp; English.
 XX
 CC The invention relates to a secreted polypeptide comprising hepatitis C
 CC virus (HCV) E2 polypeptide lacking all or a portion of its membrane
 CC spanning domain so that the E2 polypeptide is capable of secretion into
 CC growth medium when expressed recombinantly in a host cell. The
 CC polypeptide may also lack a portion of its C-terminus. The HCV E2
 CC secreted polypeptide is useful for producing anti-HCV antibodies. A
 CC purified immunogenic polypeptide comprising HCV E2 is useful for treating
 CC HCV infection and for providing immune protection against HCV infection
 CC by administering it to a subject having or at risk of having HCV
 CC infection or in need of protection. The method further comprises
 CC administering an immunomodulatory agent such as interleukin-2 (IL-2),
 CC granulocyte macrophage-colony stimulating factor (GM-CSF) or
 CC gamma-interferon. The polypeptide is useful as a vaccine, and with other
 CC HCV proteins to form a multi-component HCV vaccine for prophylactic or
 CC therapeutic treatment of HCV infection. This sequence represents an N-
 CC and C-terminally truncated HCV delta-delta E2 genotype 1a protein.

```
XX SQ Sequence 250 AA;
Query Match 100.0%; Score 20; DB 23; Length 250;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTKVCQA 20
  |||||
Db 136 GNWFGCTWNNSTGFTKVCQA 155

RESULT 11
AAB68043
ID AAB68043 standard; protein; 254 AA.
XX
AC AAB68043;
XX
DT 29-JUN-2001 (first entry)
XX
DE Amino acid sequence of water soluble variant of envelope E2 protein.
XX
KW E2 protein; HCV; envelope protein; E2 protein; HCV infection;
KW HCV attachment.
XX
OS Synthetic.
OS Hepatitis C virus.
XX
PN WO200122984-A1.
XX
PD 05-APR-2001.
XX
PF 26-SEP-2000; 2000WO-US26395.
XX
PR 29-SEP-1999; 99US-0407430.
XX
PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI Worman HJ, Mamiya N;
XX
DR WPI; 2001-273486/28.
XX
PT Treating or preventing hepatitis C virus infection in a subject,
PT involves administering hepatitis C virus envelope protein E2 binding
PT agents -
XX
PS Claim 5; Fig 8; 46pp; English.
XX
CC The present sequence represents a water soluble variant of a Hepatitis C
CC virus (HCV) envelope E2 protein. Agents that bind to the HCV E2 protein
CC (such as the human E2 protein), and so inhibit the attachment of HCV onto
CC cells (especially liver cells), are used to treat HCV infections in
CC mammals, in particular humans. The specification also describes a method
CC for identifying a compound which can be used for treating or preventing
CC HCV in a subject and which can inhibit the attachment of HCV onto cells
CC by inhibiting the binding of HCV envelope E2 protein to a cellular
CC protein associated with HCV attachment and entry into cells. The method
CC comprises incubating the compound, HCV envelope E2 protein or its variant
CC and a cellular protein capable of specifically binding to the HCV E2
CC protein under suitable reaction conditions; determining the interactions
CC between HCV envelope E2 protein and cellular protein in the presence and
CC absence of the compound; and comparing the interaction to identify a
CC compound which can inhibit the attachment of HCV onto cells.
XX
SQ Sequence 254 AA;
Query Match 100.0%; Score 20; DB 22; Length 254;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTKVCQA 20
  |||||
Db 142 GNWFGCTWNNSTGFTKVCQA 161

RESULT 12
AAE00445
ID AAE00445 standard; Protein; 271 AA.
XX
AC AAE00445;
XX
DT 19-JUN-2001 (first entry)
XX
DE HCV E2-634 HVR1 construct containing E1 signal sequence and truncated E2.
XX
KW Hepatitis C virus; HCV; hypervariable region one; HVR1; vaccine;
KW antiviral; gene therapy; envelope 2 protein; E2; immunisation;
KW HCV infection; viral replication; passive immunoprophylaxis.
XX
OS Hepatitis C virus.
OS Synthetbc.
XX
FH Key Location/Qualifiers
FT Peptide 1..20
FT /label= Signal_peptide
FT /note= "Derived from endoplasmic reticulum E1
FT signal sequence (364-383 amino acids)"
FT Protein 21..271
FT /label= Mature_C_terminal_truncated_E2_protein
XX
PN WO200121807-A1.
XX
PD 29-MAR-2001.
XX
PF 22-SEP-2000; 2000WO-US25987.
XX
PR 23-SEP-1999; 99US-0155823.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Fornis X, Bukh J, Emerson SU, Purcell RH;
XX
DR WPI; 2001-266076/27.
XX
PT Novel nucleic acid molecules that encode hepatitis C virus envelope 2
PT protein lacking all or part of hypervariable region 1 of envelope
PT protein, useful as vaccine components for treating or preventing HCV
PT infections -
XX
PS Example; Page -; 80pp; English.
XX
CC The present sequence is hepatitis C virus (HCV) E2-634 HVR1 construct
CC containing the endoplasmic reticulum signal sequence of envelope protein
CC E1 and carboxy-terminal truncated E2 protein lacking the hypervariable
CC region one (HVR1). The HCV E2 protein lacking HVR1 DNA is useful for
CC producing infectious HCV and chimeric HCV viruses which are useful for
CC for identifying cell lines capable of replicating the replication of
CC viruses. The infectious HCV and HVR1-chimeric HCV are used in the
CC production of attenuated or inactivated vaccines which are useful for
CC treating or preventing HCV in a mammal by immunisation. The host cells
CC expressing the H77C(HVR1) DNA is useful as an immunogen to stimulate a
CC protective immune response to HCV. The immunogens are useful for
CC producing protective antibodies to HCV. The antibodies produced are used
CC in passive immunoprophylaxis for treatment of diseases caused by HCV in
CC animals, especially humans. The H77C(HVR1) DNA is also useful in gene
CC therapy.
CC Note: The present sequence is not shown in the specification but is
CC derived from Hepatitis C virus envelope 2 protein lacking hypervariable
CC region 1 (AAE00442) referred as SEQ ID NO: 2 and shown in figure 1.
XX
SQ Sequence 271 AA;
Query Match 100.0%; Score 20; DB 22; Length 271;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



```
QY 1 GNWFGCTWNNSTGFTKVCGA 20
Db 157 GNWFGCTWNNSTGFTKVCGA 176

RESULT 13
AAE00446
XX AC AAE00446;
XX XX
XX 19-JUN-2001 (first entry)
XX HCV E2-634 HVR1-mut containing E1 signal sequence and truncated E2.
XX Hepatitis C virus; HCV; hypervariable region one; HVR1; vaccine;
KW antiviral; gene therapy; envelope 2 protein; E2; immunisation; mutein;
KW HCV infection; viral replication; passive immunoprophylaxis; mutant.
XX Hepatitis C virus.
OS Synthetic.
XX XX
XX Key Location/Qualifiers
XX Peptide 1..20
XX /label= Signal_peptide
XX /note= "Derived from endoplasmic reticulum E1
XX signal sequence (364-383 amino acids)"
XX Protein 21..271
XX /label= Mature_C_terminal_truncated_E2_protein
XX Misc-difference 225
XX /note= "Wild type Leu substituted with His"
XX XX
XX WO200121807-A1.
XX PN
XX 29-MAR-2001.
XX PD
XX 22-SEP-2000; 2000WO-US25987.
XX PF
XX 23-SEP-1999; 99US-0155823.
XX PR
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PA
XX Forns X, Bukh J, Emerson SU, Purcell RH;
XX WPI; 2001-266076/27.
XX DR
XX Novel nucleic acid molecules that encode hepatitis C virus envelope 2
XX protein lacking all or part of hypervariable region 1 of envelope
XX infections -
XX XX
XX Example; Page -; 80pp; English.
XX CC
XX The present sequence is hepatitis C virus (HCV) E2-634 HVR1 construct
XX containing the endoplasmic reticulum signal sequence of envelope protein
XX E1 and carboxy-terminal truncated E2 protein lacking the hypervariable
XX region one (HVR1) with a replacement of leucine for histidine. The HCV
XX E2 protein lacking HVR1 DNA is useful for producing infectious HCV and
XX chimeric HCV viruses which are useful for identifying cell lines
XX capable of supporting the replication of viruses. The infectious HCV and
XX HVR1-chimeric HCV are used in the production of attenuated or inactivated
XX vaccines which are useful for treating or preventing HCV in a mammal by
XX immunisation. The host cells expressing the H77C(HVR1) DNA is useful as
XX an immunogen to stimulate a protective immune response to HCV. The
XX immunogens are useful for producing protective antibodies to HCV. The
XX antibodies produced are used in passive immunoprophylaxis for treatment
XX of diseases caused by HCV in animals, especially humans. The H77C(HVR1)
XX DNA is also useful in gene therapy.
XX Note: The present sequence is not shown in the specification but is
XX derived from Hepatitis C virus envelope 2 protein lacking hypervariable
XX region 1 (AAE00442) referred as SEQ ID NO: 2 and shown in figure 1.
XX SQ Sequence 271 AA;

Query Match 100.0%; Score 20; DB 14; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTKVCGA 20
Db 157 GNWFGCTWNNSTGFTKVCGA 176

RESULT 14
AAR33997 0
XX ID AAR33997 standard; Protein; 278 AA.
XX AC AAR33997;
XX XX
XX 25-MAR-2003 (updated)
XX 26-JUL-1993 (first entry)
XX DT
XX Th E2/NS1 protein.
XX DE
XX Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
KW domain; immunological; cross-reactive; envelope protein; vaccine;
KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.
XX OS Synthetic.
XX WO9306126-A1.
XX PN
XX 01-APR-1993.
XX PD
XX 11-SEP-1992; 92WO-US07683.
XX PF
XX 13-SEP-1991; 91US-0759575.
XX PR
XX (CHIR ) CHIRON CORP.
XX PA
XX Houghton M, Weiner AJ;
XX PI
XX WPI; 1993-117468/14.
XX DR
XX Immuno-reactive hepatitis C virus polypeptide compans. - contg.
XX at least 2 sequences from the first variable domain of distinct
XX HCV isolates
XX PS Disclosure; Fig 3; 106pp; English.
XX CC
XX The sequences given in AAR33992-002 represent a portion of the E2/NS1
XX protein encoded by group I and group II HCV isolates, from amino acid
XX 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
XX interest. E2/NS1 contains an N-terminal hypervariable domain of about
XX 30 amino acids which shows large variation between nearly all isolates.
XX This is an important immunoreactive domain. This putative envelope
XX glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera
XX virus) envelope polypeptide of the pestiviruses and the NS1 of the
XX flaviviruses, both of which confer protective immunity in hosts
XX vaccinated with these polypeptides. It has been discovered that a
XX number of important HCV epitopes vary among viral isolates and that
XX these epitopes can be mapped to specific domains. This meant that
XX immunologically cross-reactive polypeptides which focus on variable
XX rather than constant domains can be produced. See also AAQ39134-48
XX and AAR33982-91.
XX (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 278 AA;

Query Match 100.0%; Score 20; DB 14; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTKVCGA 20
Db 157 GNWFGCTWNNSTGFTKVCGA 176
```

Qy 1 GNWFGCTWMNSTGFTKVCGA 20
| | | | | | | | | | | | | | | | | |
Db 188 GNWFGCTWMNSTGFTKVCGA 207

Search completed: November 21, 2003, 20:58:05
.Job time : 32.15 secs

Db 178 GNWFGCTWMNSTGFTKVCGA 197

RESULT 15
AAW00929
ID AAW00929 standard; Protein; 305 AA.
XX
AC AAW00929;
XX
DT 04-NOV-1997 (first entry)
XX
DE Recombinant HCV E2 antigen.
XX
KW HCV; E2; antigen; non-secretor gene; protein secretion; vaccine;
KW plasmid 577.
XX
OS Chimeric hepatitis C virus;
XX Chimeric synthetic.
FH Key Location/Qualifiers
FT Peptide 1..19
FT /label= Sig_peptide
FT Cleavage-site 20
FT /note= "mammalian secretion signal
FT peptidase cleavage site"
FT Protein 20..305
FT /label= Mat_protein
FT Peptide 25..28
FT /note= "N-terminal sequence of prourokinase"
XX
PN WO9641179-A1.
XX
PD 19-DEC-1996.
XX
PF 05-JUN-1996; 96WO-US09345.
XX
PR 07-JUN-1995; 95US-0478073.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Lesniewski RR, Okasinski GF, Schaefer VG, Suhar TS;
XX
DR WPI: 1997-108653/10.
DR N-PSDB; AAT13899.
XX
PT New expression system for proteins, partic. HCV antigens - for use
PT in assays for screening and prognostic applications and for use in
PT vaccines
XX
PS Example 1; Page 29-30; 40pp; English.
XX
CC This sequence comprises the conceptual translation product of
CC a hepatitis C virus E2 antigen expression cassette (AAT13899). The
CC encoded E2 antigen is truncated at amino acid residue 644 of HCV,
CC and contains an N-terminal sequence (SHEL) from human prourokinase
CC intended to promote signal protease processing, efficient secretion
CC and final product stability. Glycosylated E2 can be expressed in
CC mammalian host cells utilising claimed plasmid 577. Fusion
CC proteins produced by plasmid 577 are used in claimed assays and
CC test kits for detecting anti-analyte antibody as well as in claimed
CC vaccines for treatment of infection. The plasmid can be used to
CC produce high levels of proteins that would not normally be
CC expressed in mammalian cells due to the non-secretory nature of the
CC gene. The system is esp. used for expressing HCV proteins, allowing
CC proper processing, glycosylation and conformation of the viral
CC protein.
XX
SQ Sequence 305 AA;

Query Match 100.0%; Score 20; DB 18; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.6 Seconds
(without alignments)
79.832 Million cell updates/sec

Title: US-09-973-025-83
Perfect score: 20
Sequence: 1 GNWFGCTWMNSTGFTKVCGA 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0
Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pcp.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pcp.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pcp.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pcp.*
5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pcp.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-08-612-973-83
2	20	100.0	20	3	US-08-927-597-83
3	20	100.0	20	4	US-08-635-886C-39
4	20	100.0	26	1	US-08-262-037-2
5	20	100.0	34	1	US-08-262-037-41
6	20	100.0	42	1	US-08-262-037-42
7	20	100.0	50	1	US-08-262-037-43
8	20	100.0	179	3	US-08-444-818-77
9	20	100.0	278	1	US-08-440-103-15
10	20	100.0	278	1	US-08-440-542-15
11	20	100.0	278	1	US-08-231-368-15
12	20	100.0	278	1	US-08-440-210-15
13	20	100.0	278	4	US-09-046-604-15
14	20	100.0	305	3	US-08-478-073-2
15	20	100.0	333	1	US-08-453-552-12
16	20	100.0	333	2	US-08-710-637-12
17	20	100.0	333	5	PCT-US93-00907-12
18	20	100.0	337	1	US-08-188-281B-7
19	20	100.0	337	5	PCT-US94-07280-7
20	20	100.0	337	5	PCT-US95-01087-7
21	20	100.0	367	1	US-08-188-281B-9
22	20	100.0	367	1	US-08-453-552-6
23	20	100.0	367	2	US-08-710-637-6
24	20	100.0	367	5	PCT-US93-00907-6
25	20	100.0	367	5	PCT-US94-07280-9
26	20	100.0	367	5	PCT-US95-01087-9
27	20	100.0	377	1	US-08-188-281B-17

28	20	100.0	377	5	PCT-US94-07280-17	Sequence 17, Appl
29	20	100.0	377	5	PCT-US95-01087-17	Sequence 17, Appl
30	20	100.0	397	1	US-08-188-281B-11	Sequence 11, Appl
31	20	100.0	397	5	PCT-US94-07280-11	Sequence 11, Appl
32	20	100.0	397	5	PCT-US95-01087-11	Sequence 11, Appl
33	20	100.0	399	1	US-08-453-552-10	Sequence 10, Appl
34	20	100.0	399	2	US-08-710-637-10	Sequence 10, Appl
35	20	100.0	399	5	PCT-US93-00907-10	Sequence 10, Appl
36	20	100.0	402	1	US-08-460-806-13	Sequence 13, Appl
37	20	100.0	402	1	US-08-460-806-15	Sequence 15, Appl
38	20	100.0	402	1	US-08-460-806-17	Sequence 17, Appl
39	20	100.0	402	1	US-08-325-630-13	Sequence 13, Appl
40	20	100.0	402	1	US-08-325-630-15	Sequence 15, Appl
41	20	100.0	402	1	US-08-325-630-17	Sequence 17, Appl
42	20	100.0	403	2	US-08-483-695-39	Sequence 39, Appl
43	20	100.0	403	2	US-07-965-285-39	Sequence 39, Appl
44	20	100.0	403	2	US-08-487-231-39	Sequence 39, Appl
45	20	100.0	403	3	US-09-201-912-39	Sequence 39, Appl

ALIGNMENTS

RESULT 1
US-08-612-973-83
; Sequence 83, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-612-973-83

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.6e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GNWFGCTWMNSTGFTKVCGA 20
|||||

Db 1 GNWFGCTWMNSTGFTKVCGA 20

RESULT 2

US-08-927-597-83
; Sequence 83, Application US/08927597
; Patent No. 6245503

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT
APPLICANT: BOSMAN, FONS
APPLICANT: DE MARTYNOFF, GUY
APPLICANT: BUYSSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHUYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/927,597

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 83:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-927-597-83

Query Match 100.0%; Score 20; DB 3; Length 20;

Best Local Similarity 100.0%; Pred. No. 8.6e-15;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GNWFGCTWMNSTGFTKVCGA 20

Db 1 GNWFGCTWMNSTGFTKVCGA 20

RESULT 3

US-08-635-886C-39
; Sequence 39, Application US/08635886C
; Patent No. 655114

GENERAL INFORMATION:

APPLICANT: LEROUX-ROELS, Geert

APPLICANT: DELEYS, Robert

APPLICANT: MAERTENS, Geert

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C

TITLE OF INVENTION: VIRUS

FILE REFERENCE: 2752-18

CURRENT APPLICATION NUMBER: US/08/635,886C

CURRENT FILING DATE: 1996-04-25

PRIOR APPLICATION NUMBER: PCT/EP94/03555

0
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6

; PRIOR FILING DATE: 1993-11-04

; NUMBER OF SEQ ID NOS: 286

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 39

; LENGTH: 20

; TYPE: PRT

; ORGANISM: hepatitis C virus

US-08-635-886C-39

Query Match 100.0%; Score 20; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 8.6e-15;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GNWFGCTWMNSTGFTKVCGA 20

Db 1 GNWFGCTWMNSTGFTKVCGA 20

RESULT 4

US-08-262-037-2

; Sequence 2, Application US/08262037

; Patent No. 5747239

; GENERAL INFORMATION:

APPLICANT: Chang Yi Wang and Barbara Hosein

TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR

TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV

TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES

NUMBER OF SEQUENCES: 136

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVE.

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/262,037

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/719,819

FILING DATE: 24-June-1991

APPLICATION NUMBER: 07/667,275

FILING DATE: 11-Mar-1991

APPLICATION NUMBER: 07/651,735

FILING DATE: 07-Feb-1991

APPLICATION NUMBER: 07/558,799

FILING DATE: 26-July-1990

APPLICATION NUMBER: 07/510,153

FILING DATE: 16-April-1990

ATTORNEY/AGENT INFORMATION:

NAME: Maria C. H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4043 US3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-758-4800

TELEFAX: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 26 amino acids

TYPE: Amino acid

STRANDEDNESS:

TOPOLOGY: Unknown

US-08-262-037-2

```
Query Match 100.0%; Score 20; DB 1; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNNFGCTWNNSTGFTKVCGA 20
Db 4 GNNFGCTWNNSTGFTKVCGA 23

RESULT 5
US-08-262-037-41
; Sequence 41, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-262-037-41

Query Match 100.0%; Score 20; DB 1; Length 34;
Best Local Similarity 100.0%; Pred. No. 1.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNNFGCTWNNSTGFTKVCGA 20
Db 12 GNNFGCTWNNSTGFTKVCGA 31

RESULT 6
US-08-262-037-42
; Sequence 42, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-262-037-42

Query Match 100.0%; Score 20; DB 1; Length 42;
Best Local Similarity 100.0%; Pred. No. 1.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNNFGCTWNNSTGFTKVCGA 20
Db 20 GNNFGCTWNNSTGFTKVCGA 39

RESULT 7
US-08-262-037-43
; Sequence 43, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-262-037-43
```

CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/262,037
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/719,819
FILING DATE: 24-June-1991
APPLICATION NUMBER: 07/667,275
FILING DATE: 11-Mar-1991
APPLICATION NUMBER: 07/651,735
FILING DATE: 07-Feb-1991
APPLICATION NUMBER: 07/558,799
FILING DATE: 26-July-1990
APPLICATION NUMBER: 07/510,153
FILING DATE: 16-April-1990
ATTORNEY/AGENT INFORMATION:
NAME: Maria C. H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4043 US3
TELEPHONE: 212-758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 amino acids
TYPE: Amino acid
STRANDEDNESS:
TOPOLOGY: Unknown
US-08-262-037-43

Query Match 100.0%; Score 20; DB 1; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTKVCGA 20
DB 28 GNWFGCTWNNSTGFTKVCGA 47

RESULT 8
US-08-444-818-77
Sequence 77, Application US/08444818
Patent No. 6150087
GENERAL INFORMATION:
APPLICANT: Chien, David Y.
ATTORNEY/AGENT INFORMATION:
TITLE OF INVENTION: NANEV Diagnostics and Vaccines
NUMBER OF SEQUENCES: 777
CORRESPONDENCE ADDRESS:
ADDRESS: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,818
FILING DATE:

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/403,590
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Hardin, Ailsa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0110.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508)359-3876
TELEFAX: (508)359-3885
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 179 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-444-818-77
Query Match 100.0%; Score 20; DB 3; Length 179;
Best Local Similarity 100.0%; Pred. No. 4.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GNWFGCTWNNSTGFTKVCGA 20
DB 37 GNWFGCTWNNSTGFTKVCGA 56
RESULT 9
US-08-440-103-15
Sequence 15, Application US/08440103
Patent No. 5670152
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,103
FILING DATE: 12-MAY-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 278 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-103-15

Query Match 100.0%; Score 20; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 6.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTKVCGA 20
DB 178 GNWFGCTWNNSTGFTKVCGA 197

RESULT 10
US-08-440-542-15
; Sequence 15, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,542
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-542-15

Query Match 100.0%; Score 20; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 6.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTKVCGA 20
DB 178 GNWFGCTWNNSTGFTKVCGA 197

RESULT 11
US-08-231-368-15
; Sequence 15, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions

; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-231-368-15

Query Match 100.0%; Score 20; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 6.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTKVCGA 20
DB 178 GNWFGCTWNNSTGFTKVCGA 197

RESULT 12
US-08-440-210-15
; Sequence 15, Application US/08440210
; Patent No. 5766845
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,210
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:

APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 278 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-210-15

Query Match 100.0%; Score 20; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 6.8e-14; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;
QY 1 GNWFGCTWMNSTGFTKVCGA 20
Db 178 GNWFGCTWMNSTGFTKVCGA 197

RESULT 13
US-09-046-604-15
Sequence 15, Application US/09046604
Patent No. 6303292
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/046,604
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 278 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-046-604-15

Query Match 100.0%; Score 20; DB 4; Length 278;
Best Local Similarity 100.0%; Pred. No. 6.8e-14; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;
QY 1 GNWFGCTWMNSTGFTKVCGA 20
Db 178 GNWFGCTWMNSTGFTKVCGA 197

RESULT 14
US-08-478-073-2
Sequence 2, Application US/08478073
Patent No. 6020122
GENERAL INFORMATION:
APPLICANT: Okasinski, Gregory F.
APPLICANT: Schaefer, Verlyn G.
APPLICANT: Suhar, Thomas S.
APPLICANT: Lesniewski, Richard R.
TITLE OF INVENTION: EXPRESSION SYSTEM FOR NON-SECRETOR GENES
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES
STREET: ONE HUNDRED ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,073
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Porembski, Priscilla E.
REGISTRATION NUMBER: 33207
REFERENCE/DOCKET NUMBER: 5763.US.01
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-073-2

Query Match 100.0%; Score 20; DB 3; Length 305;
Best Local Similarity 100.0%; Pred. No. 7.3e-14; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;
QY 1 GNWFGCTWMNSTGFTKVCGA 20
Db 188 GNWFGCTWMNSTGFTKVCGA 207

RESULT 15
US-08-453-552-12
Sequence 12, Application US/08453552
Patent No. 5667992
GENERAL INFORMATION:
APPLICANT: CASEY, JAMES M.
APPLICANT: BODE, SUZANNE L.
APPLICANT: ZECK, BILLY J.
APPLICANT: YAMAGUCHI, JULIE
APPLICANT: FRAIL, DONALD B.
APPLICANT: DESAI, SURESH M.
APPLICANT: DEVARE, SUSHIL G.
TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
MOLECULE TYPE: PROTEINS
NUMBER OF SEQUENCES: 12

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
;; STREET: ONE ABBOTT PARK ROAD
;; CITY: ABBOTT PARK
;; STATE: IL
;; COUNTRY: USA
;; ZIP: 60064-3500
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/453,552
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: FOREMSKI, PRISCILLA E.
;; REGISTRATION NUMBER: 33,207
;; REFERENCE/DOCKET NUMBER: 5131.US.D1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 708-937-6365
;; TELEFAX: 708-937-9556
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 333 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-453-552-12

Query Match 100.0%; Score 20; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 7.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GNWFGCTWNNSTGFTKVCGA 20
|||
Db 196 GNWFGCTWNNSTGFTKVCGA 215

Search completed: November 21, 2003, 21:15:16
Job time : 10.6 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time 19.5 Seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-83

Perfect score: 20

Sequence: 1 GWFCTWNNSTGTFKVCGA 20

Scoring table:

OLIGO 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	10	US-09-973-025-83
2	20	100.0	20	11	US-09-899-303-83
3	20	100.0	20	11	US-09-995-808-83
4	20	100.0	20	11	US-09-995-860-83
5	20	100.0	20	12	US-09-995-791-83
6	20	100.0	96	10	US-09-921-397-82
7	20	100.0	176	10	US-09-921-397-81
8	20	100.0	250	10	US-09-952-572-8
9	20	100.0	254	10	US-09-407-430-3
10	20	100.0	350	10	US-09-929-955-4
11	20	100.0	350	14	US-10-104-966-4
12	20	100.0	363	10	US-09-407-430-2
13	20	100.0	363	12	US-10-128-587A-97
14	20	100.0	363	15	US-10-128-590-97
15	20	100.0	637	12	US-10-187-257-4

16	20	100.0	637	12	US-10-265-083-2
17	20	100.0	2894	10	US-09-941-611-23
18	20	100.0	2894	15	US-10-044-995-23
19	20	100.0	3011	9	US-09-742-659-4
20	20	100.0	3011	9	US-09-916-359-2
21	20	100.0	3011	10	US-09-238-076-20
22	20	100.0	3011	10	US-09-952-572-9
23	20	100.0	3011	10	US-09-929-955-1
24	20	100.0	3011	10	US-09-747-419-20
25	20	100.0	3011	11	US-09-891-894-3
26	20	100.0	3011	11	US-09-995-937-20
27	20	100.0	3011	11	US-09-917-563-20
28	20	100.0	3011	12	US-10-184-150-3
29	20	100.0	3011	14	US-10-104-966-1
30	20	100.0	3011	15	US-10-259-275-20
31	20	100.0	3011	16	US-10-232-643-6
32	20	100.0	3012	10	US-09-238-076-2
33	20	100.0	3012	11	US-09-995-937-2
34	20	100.0	3012	11	US-09-917-563-2
35	16	80.0	347	11	US-09-194-949-9
36	16	80.0	539	11	US-09-194-949-11
37	16	80.0	2985	15	US-10-259-275-40
38	10	50.0	44	15	US-10-318-200-14
39	10	50.0	290	12	US-10-128-587A-3
40	10	50.0	290	15	US-10-128-590-3
41	10	50.0	301	12	US-10-128-587A-5
42	10	50.0	301	15	US-10-128-590-5
43	10	50.0	314	10	US-09-973-025-42
44	10	50.0	314	11	US-09-899-303-42
45	10	50.0	314	11	US-09-995-808-42

ALIGNMENTS

RESULT 1

US-09-973-025-83
; Sequence 83, Application US/09973025
; Publication No. US20020182706A1

; GENERAL INFORMATION:

; APPLICANT: MAERTENS, GEERT

; BOSMAN, FONS

; DE MARTYNOFF, GUY

; BUYSE, MARIE-ANGE

; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

; NUMBER OF SEQUENCES: 111

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: NIXON & VANDERHYE P.C.

; STREET: 1100 NORTH GLEBE ROAD

; CITY: ARLINGTON

; STATE: VIRGINIA

; COUNTRY: U.S.A.

; ZIP: 22201-4714

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/973,025

; FILING DATE: 10-Oct-2001

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/612,973

; FILING DATE: 11-MAR-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: BYRNE, THOMAS E.

; REGISTRATION NUMBER: 32,205

; REFERENCE/DOCKET NUMBER: 1487-10

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 816-4000

; TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 83:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 83:
US-09-973-025-83

Query Match 100.0%; Score 20; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.5e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||||

Db 1 GNWFGCTWMNSTGFTKVCGA 20
|||||

RESULT 2
US-09-899-303-83
; Sequence 83, Application US/09899303
; Publication No. US20030036110A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION NUMBER: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 83:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 83:
US-09-899-303-83

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.5e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||||

Db 1 GNWFGCTWMNSTGFTKVCGA 20
|||||

RESULT 3
US-09-995-808-83
; Sequence 83, Application US/0995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 83
; LENGTH: 20
; TYPE: PRP
; ORGANISM: Hepatitis C virus
US-09-995-808-83

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.5e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||||

Db 1 GNWFGCTWMNSTGFTKVCGA 20
|||||

RESULT 4
US-09-995-860-83
; Sequence 83, Application US/0995860
; Publication No. US20030118603A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and therapeutic use.
; FILE REFERENCE: 2551-89
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 83
; LENGTH: 20
; TYPE: PRP
; ORGANISM: Hepatitis C virus
US-09-995-860-83

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.5e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||||

Db 1 GNWFGCTWMNSTGFTKVCGA 20

RESULT 3
US-09-995-808-83
; Sequence 83, Application US/0995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 83
; LENGTH: 20
; TYPE: PRP
; ORGANISM: Hepatitis C virus
US-09-995-808-83

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.5e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||||

Db 1 GNWFGCTWMNSTGFTKVCGA 20
|||||

RESULT 4
US-09-995-860-83
; Sequence 83, Application US/0995860
; Publication No. US20030118603A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and therapeutic use.
; FILE REFERENCE: 2551-89
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 83
; LENGTH: 20
; TYPE: PRP
; ORGANISM: Hepatitis C virus
US-09-995-860-83

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.5e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||||

Db 1 GNWFGCTWMNSTGFTKVCGA 20
|||||

RESULT 5
US-09-995-791-83
; Sequence 83, Application US/0995791
; Publication No. US20030147918A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and therapeutic use.
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 83

```
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-83

Query Match      100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.5e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNNFGCTWNNSTGFTKVCGA 20
   |||||
Db 1 GNNFGCTWNNSTGFTKVCGA 20

RESULT 6
US-09-921-397-82
; Sequence 82, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; FILE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 82
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-82

Query Match      100.0%; Score 20; DB 10; Length 96;
Best Local Similarity 100.0%; Pred. No. 2.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNNFGCTWNNSTGFTKVCGA 20
   |||||
Db 61 GNNFGCTWNNSTGFTKVCGA 80

RESULT 7
US-09-921-397-81
; Sequence 81, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; FILE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 81
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-81

Query Match      100.0%; Score 20; DB 10; Length 176;
Best Local Similarity 100.0%; Pred. No. 3.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNNFGCTWNNSTGFTKVCGA 20
   |||||

; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-83

Query Match      100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.5e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNNFGCTWNNSTGFTKVCGA 20
   |||||
Db 1 GNNFGCTWNNSTGFTKVCGA 20

RESULT 8
US-09-952-572-8
; Sequence 8, Application US/09952572
; Patent No. US20020119495A1
; GENERAL INFORMATION:
; APPLICANT: HAWAII BIOTECHNOLOGY GROUP, Inc.
; APPLICANT: NAKANO, Eileen
; APPLICANT: CLEMENTS, David
; APPLICANT: HUMPHREYS, Tom
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITION OF HEPATITIS C
; FILE REFERENCE: HAWBIO1100
; CURRENT APPLICATION NUMBER: US/09/952,572
; CURRENT FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: US 60/230,927
; PRIOR FILING DATE: 2000-09-13
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-09-952-572-8

Query Match      100.0%; Score 20; DB 10; Length 250;
Best Local Similarity 100.0%; Pred. No. 5.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNNFGCTWNNSTGFTKVCGA 20
   |||||
Db 136 GNNFGCTWNNSTGFTKVCGA 155

RESULT 9
US-09-407-430-3
; Sequence 3, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Norman, Howard J.
; APPLICANT: Mamlya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; FILE OF INVENTION: HEPATITIS C VIRUS INFECTION
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-407-430-3

Query Match      100.0%; Score 20; DB 10; Length 254;
Best Local Similarity 100.0%; Pred. No. 5.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNNFGCTWNNSTGFTKVCGA 20
   |||||
Db 142 GNNFGCTWNNSTGFTKVCGA 161

RESULT 10
US-09-929-955-4
; Sequence 4, Application US/09929955
; Patent No. US20020136740A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; FILE OF INVENTION: METHODS OF USE THEREOF
```

; FILE REFERENCE: TRIPEP.23AUS2
; CURRENT APPLICATION NUMBER: US/09/929,955
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/225,767
; PRIOR FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-09-929-955-4

Query Match 100.0%; Score 20; DB 10; Length 350;
Best Local Similarity 100.0%; Pred. No. 6.7e-14; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||||
DB 168 GNWFGCTWMNSTGFTKVCGA 187

RESULT 11
US-10-104-966-4
; Sequence 4, Application US/10104966
; Publication No. US20020155124A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; FILE REFERENCE: TRIPEP.23AUSC1
; CURRENT APPLICATION NUMBER: US/10/104,966
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-10-104-966-4

Query Match 100.0%; Score 20; DB 14; Length 350;
Best Local Similarity 100.0%; Pred. No. 6.7e-14; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||||
DB 168 GNWFGCTWMNSTGFTKVCGA 187

RESULT 12
US-09-407-430-2
; Sequence 2, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoko
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; FILE REFERENCE: 0575/54805

; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-09-407-430-2

Query Match 100.0%; Score 20; DB 10; Length 363;
Best Local Similarity 100.0%; Pred. No. 6.9e-14; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||||
DB 164 GNWFGCTWMNSTGFTKVCGA 183

RESULT 13
US-10-128-587A-97
; Sequence 97, Application US/10128587A
; Publication No. US20030152940A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Constructs and methods for expression of recombinant HCV envelope
; FILE REFERENCE: 134 PCT
; CURRENT APPLICATION NUMBER: US/10/128,587A
; CURRENT FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: hepatitis C virus
US-10-128-587A-97

Query Match 100.0%; Score 20; DB 12; Length 363;
Best Local Similarity 100.0%; Pred. No. 6.9e-14; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||||
DB 164 GNWFGCTWMNSTGFTKVCGA 183

RESULT 14
US-10-128-590-97
; Sequence 97, Application US/10128590
; Publication No. US20030108561A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Core-glycosylated HCV envelope proteins
; FILE REFERENCE: 135 PCT
; CURRENT APPLICATION NUMBER: US/10/128,590
; CURRENT FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-128-590-97

Query Match 100.0%; Score 20; DB 15; Length 363;
Best Local Similarity 100.0%; Pred. No. 6.9e-14; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||||

Db 164 GNWFGCTWNNSTGFTKVCGA 183

RESULT 15

US-10-187-257-4
; Sequence 4, Application US/10187257
; Publication NO. US20030138458A1
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: O'HAGAN, Derek
; TITLE OF INVENTION: HCV E1E2 VACCINE COMPOSITIONS
; FILE REFERENCE: 2302-17206
; CURRENT APPLICATION NUMBER: US/10/187,257
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
; OTHER INFORMATION: amino acid
US-10-187-257-4

Query Match 100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTKVCGA 20
| | | | | | | | | | | | | | | | | | | | | |
Db 375 GNWFGCTWNNSTGFTKVCGA 394

Search completed: November 21, 2003, 22:19:39
Job time : 19.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 163.1 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-83
Perfect score: 20
Sequence: 1 GNWFGCTWNNSTGTTCVCA 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0
Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents AA Main:*

1:	/cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
2:	/cgn2_6/ptodata/1/paa/US06_COMB.pep.*
3:	/cgn2_6/ptodata/1/paa/US07_COMB.pep.*
4:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*
5:	/cgn2_6/ptodata/1/paa/US081_COMB.pep.*
6:	/cgn2_6/ptodata/1/paa/US082_COMB.pep.*
7:	/cgn2_6/ptodata/1/paa/US083_COMB.pep.*
8:	/cgn2_6/ptodata/1/paa/US084_COMB.pep.*
9:	/cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10:	/cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11:	/cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12:	/cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13:	/cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14:	/cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15:	/cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16:	/cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17:	/cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18:	/cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19:	/cgn2_6/ptodata/1/paa/US095_COMB.pep.*
20:	/cgn2_6/ptodata/1/paa/US096_COMB.pep.*
21:	/cgn2_6/ptodata/1/paa/US097A_COMB.pep.*
22:	/cgn2_6/ptodata/1/paa/US097B_COMB.pep.*
23:	/cgn2_6/ptodata/1/paa/US098_COMB.pep.*
24:	/cgn2_6/ptodata/1/paa/US099A_COMB.pep.*
25:	/cgn2_6/ptodata/1/paa/US099B_COMB.pep.*
26:	/cgn2_6/ptodata/1/paa/US100_COMB.pep.*
27:	/cgn2_6/ptodata/1/paa/US101_COMB.pep.*
28:	/cgn2_6/ptodata/1/paa/US102_COMB.pep.*
29:	/cgn2_6/ptodata/1/paa/US103_COMB.pep.*
30:	/cgn2_6/ptodata/1/paa/US104_COMB.pep.*
31:	/cgn2_6/ptodata/1/paa/US106_COMB.pep.*
32:	/cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	20	100.0	20	13	US-08-974-685-39
					Sequence 39, Appl

2	20	100.0	20	13	US-08-974-690-39	Sequence 39, Appl
3	20	100.0	20	13	US-08-974-690A-39	Sequence 39, Appl
4	20	100.0	20	13	US-08-974-690B-39	Sequence 39, Appl
5	20	100.0	20	13	US-08-974-690C-39	Sequence 39, Appl
6	20	100.0	20	23	US-09-899-303-83	Sequence 83, Appl
7	20	100.0	20	23	US-09-899-303A-83	Sequence 83, Appl
8	20	100.0	20	25	US-09-973-025-83	Sequence 83, Appl
9	20	100.0	20	25	US-09-995-791-83	Sequence 83, Appl
10	20	100.0	20	25	US-09-995-808-83	Sequence 83, Appl
11	20	100.0	20	25	US-09-995-860-83	Sequence 83, Appl
12	20	100.0	20	26	US-10-020-510-83	Sequence 83, Appl
13	20	100.0	20	29	US-10-321-798-83	Sequence 83, Appl
14	20	100.0	26	3	US-07-667-275A-2	Sequence 2, Appl
15	20	100.0	26	8	US-08-475-482-2	Sequence 2, Appl
16	20	100.0	26	8	US-08-477-072-2	Sequence 2, Appl
17	20	100.0	26	8	US-08-477-582-2	Sequence 2, Appl
18	20	100.0	26	8	US-08-480-253-2	Sequence 2, Appl
19	20	100.0	34	8	US-08-475-482-41	Sequence 41, Appl
20	20	100.0	34	8	US-08-477-072-41	Sequence 41, Appl
21	20	100.0	34	8	US-08-477-582-41	Sequence 41, Appl
22	20	100.0	34	8	US-08-480-253-41	Sequence 41, Appl
23	20	100.0	42	8	US-08-475-482-42	Sequence 42, Appl
24	20	100.0	42	8	US-08-477-072-42	Sequence 42, Appl
25	20	100.0	42	8	US-08-477-582-42	Sequence 42, Appl
26	20	100.0	42	8	US-08-480-253-42	Sequence 42, Appl
27	20	100.0	50	8	US-08-475-482-43	Sequence 43, Appl
28	20	100.0	50	8	US-08-477-072-43	Sequence 43, Appl
29	20	100.0	50	8	US-08-477-582-43	Sequence 43, Appl
30	20	100.0	50	8	US-08-480-253-43	Sequence 43, Appl
31	20	100.0	96	24	US-09-921-397-82	Sequence 82, Appl
32	20	100.0	112	8	US-08-436-966-19	Sequence 19, Appl
33	20	100.0	176	24	US-09-921-397-81	Sequence 81, Appl
34	20	100.0	179	8	US-08-403-590B-77	Sequence 77, Appl
35	20	100.0	179	8	US-08-444-112-77	Sequence 77, Appl
36	20	100.0	250	1	PCT-US01-28767-8	Sequence 8, Appl
37	20	100.0	250	25	US-09-952-572-8	Sequence 8, Appl
38	20	100.0	254	1	PCT-US00-26395-3	Sequence 3, Appl
39	20	100.0	254	18	US-09-407-430-3	Sequence 3, Appl
40	20	100.0	278	8	US-08-471-498-15	Sequence 15, Appl
41	20	100.0	333	8	US-08-417-478-12	Sequence 12, Appl
42	20	100.0	333	8	US-08-453-613-12	Sequence 12, Appl
43	20	100.0	333	30	US-10-445-724-3	Sequence 3, Appl
44	20	100.0	333	32	US-06-409-909-4	Sequence 4, Appl
45	20	100.0	350	21	US-09-705-547-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-974-685-39
; Sequence 39, Application US/08974685
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; DELEYS, ROBERT
; MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
; THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 181
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,685
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 39:

US-08-974-685-39

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
| | | | | | | | | | | | | | | | | | | | | |
DB 1 GNWFGCTWMNSTGFTKVCGA 20

RESULT 2

US-08-974-690-39
Sequence 39, Application US/08974690
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 166

CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/635,886
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

US-08-974-690-39

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
| | | | | | | | | | | | | | | | | | | | | |
DB 1 GNWFGCTWMNSTGFTKVCGA 20

RESULT 3

US-08-974-690A-39
Sequence 39, Application US/08974690A
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690A
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 39:

SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 39:

US-08-974-690A-39

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
| | | | | | | | | | | | | | | | | | | | | |
DB 1 GNWFGCTWMNSTGFTKVCGA 20

RESULT 4

US-08-974-690B-39
Sequence 39, Application US/08974690B
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
DELEYS, ROBERT
MAERTENS, GEERT

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690B
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-08-974-690B-39

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GNWFGCTWMNSTGFTKVCGA 20
Db 1 GNWFGCTWMNSTGFTKVCGA 20

RESULT 5
US-08-974-690C-39
; Sequence 39, Application US/08974690C
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT FILING DATE: 1997-11-19
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 39
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-974-690C-39

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GNWFGCTWMNSTGFTKVCGA 20

Db 1 GNWFGCTWMNSTGFTKVCGA 20

RESULT 6

US-09-899-303-83
; Sequence 83, Application US/09899303
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 83:
US-09-899-303-83

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GNWFGCTWMNSTGFTKVCGA 20
Db 1 GNWFGCTWMNSTGFTKVCGA 20

RESULT 7

US-09-899-303A-83
; Sequence 83, Application US/09899303A
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.

STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303A
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 83:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 83:
US-09-899-303A-83

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||||
DB 1 GNWFGCTWMNSTGFTKVCGA 20
|||||

RESULT 8
US-09-973-025-83
; Sequence 83, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSER: NIXON & VANDERHUYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/973,025
FILING DATE: 10-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 83:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 83:
US-09-973-025-83

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||||
DB 1 GNWFGCTWMNSTGFTKVCGA 20
|||||

RESULT 9
US-09-995-791-83
; Sequence 83, Application US/09995791
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 83
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-83

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||||
DB 1 GNWFGCTWMNSTGFTKVCGA 20
|||||

RESULT 10
US-09-995-808-83
; Sequence 83, Application US/09995808
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 83
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-83

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTKVCGA 20
|||||
Db 1 GNWFGCTWNNSTGFTKVCGA 20

RESULT 11

US-09-995-860-83

; Sequence 83, Application US/09995860

; GENERAL INFORMATION:

; APPLICANT: Innogenetics N.V.

; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and

; FILE REFERENCE: 2551-69

; CURRENT APPLICATION NUMBER: US/09/995,860

; CURRENT FILING DATE: 2001-11-29

; NUMBER OF SEQ ID NOS: 122

; SOFTWARE: PatentIn 3.1

; SEQ ID NO 83

; LENGTH: 20

; TYPE: PRT

; ORGANISM: Hepatitis C virus

US-09-995-860-83

Query Match

Best Local Similarity 100.0%; Score 20; DB 25; Length 20;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTKVCGA 20
|||||
Db 1 GNWFGCTWNNSTGFTKVCGA 20

RESULT 12

US-10-020-510-83

; Sequence 83, Application US/10020510

; GENERAL INFORMATION:

; APPLICANT: Innogenetics N.V.

; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and

; FILE REFERENCE: 2551-72

; CURRENT APPLICATION NUMBER: US/10/020,510

; CURRENT FILING DATE: 2002-07-15

; NUMBER OF SEQ ID NOS: 122

; SOFTWARE: PatentIn 3.1

; SEQ ID NO 83

; LENGTH: 20

; TYPE: PRT

; ORGANISM: Hepatitis C virus

US-10-020-510-83

Query Match

Best Local Similarity 100.0%; Score 20; DB 26; Length 20;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTKVCGA 20
|||||
Db 1 GNWFGCTWNNSTGFTKVCGA 20

RESULT 13

US-10-321-798-83

; Sequence 83, Application US/10321798

; GENERAL INFORMATION:

; APPLICANT: Innogenetics N.V.

; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and

; FILE REFERENCE: 2551-93

; CURRENT APPLICATION NUMBER: US/10/321,798

; CURRENT FILING DATE: 2002-12-18

; PRIOR APPLICATION NUMBER: 60/418,358

; PRIOR FILING DATE: 2002-10-16

; PRIOR APPLICATION NUMBER: 10/020,510

; PRIOR FILING DATE: 2001-12-18

; NUMBER OF SEQ ID NOS: 123

; SOFTWARE: PatentIn 3.1

; SEQ ID NO 83

; LENGTH: 20

; TYPE: PRT

; ORGANISM: Hepatitis C virus

US-10-321-798-83

Query Match

Best Local Similarity 100.0%; Score 20; DB 29; Length 20;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTKVCGA 20
|||||
Db 1 GNWFGCTWNNSTGFTKVCGA 20

RESULT 14

US-07-667-275A-2

; Sequence 2, Application US/07667275A

; GENERAL INFORMATION:

; APPLICANT: Wang, Chang Yi

; APPLICANT: Hosein, Barbara M.

; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR THE

; TITLE OF INVENTION: DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV INFECTION

; TITLE OF INVENTION: AND PREVENTION THEREOF AS VACCINES

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Morgan & Finnegan

; STREET: 345 Park Avenue

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10154

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/667,275A

; FILING DATE: 19910311

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/651,735

; FILING DATE: 07-FEB-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/558,799

; FILING DATE: 26-JUL-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/510,153

; FILING DATE: 16-APR-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/481/348

; FILING DATE: 16-FEB-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Lin, Maria C.H.

; REGISTRATION NUMBER: 29,323

; REFERENCE/DOCKET NUMBER: 11514043B

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-758-4800

; TELEFAX: 212-751-6849

; TELEX: 421792

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 26 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: unknown

; MOLECULE TYPE: peptide

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

1 ORIGINAL SOURCE:
1 ORGANISM: Hepatitis C Virus
US-07-667-275A-2

Query Match 100.0%; Score 20; DB 3; Length 26;
Best Local Similarity 100.0%; Pred. No. 3.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
DB 4 GNWFGCTWMNSTGFTKVCGA 23

RESULT 15

US-08-475-482-2
; Sequence 2, Application US/08475482
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,482
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 17-June-1994
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-475-482-2

Query Match 100.0%; Score 20; DB 8; Length 26;
Best Local Similarity 100.0%; Pred. No. 3.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20

Db 4 GNWFGCTWMNSTGFTKVCGA 23
0
Search completed: November 21, 2003, 22:09:52
Job time : 163.1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.55 Seconds
(without alignments)
117.536 Million cell updates/sec

Title: US-09-973-025-83
Perfect score: 20
Sequence: 1 GNWFGCTWMNSTGFTKVCGA 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA_Nov.*
1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US16_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	20	100.0	20	6	US-10-651-165-39	Sequence 39, Appl
2	20	100.0	30	6	US-10-296-734-480	Sequence 480, App
3	20	100.0	333	1	PCT-US03-19834-3	Sequence 3, Appl
4	20	100.0	333	6	US-10-655-562-4	Sequence 4, Appl
5	20	100.0	637	1	PCT-US03-33610-4	Sequence 4, Appl
6	20	100.0	2011	6	US-10-296-734-812	Sequence 812, App
7	20	100.0	3011	1	PCT-US03-19834-2	Sequence 2, Appl
8	20	100.0	3011	6	US-10-296-734-406	Sequence 406, App
9	20	100.0	5985	6	US-10-296-734-810	Sequence 810, App
10	16	80.0	347	6	US-10-664-391-9	Sequence 9, Appl
11	16	80.0	539	6	US-10-664-391-11	Sequence 11, Appl
12	16	80.0	1026	1	PCT-US03-20409-3	Sequence 3, Appl
13	16	80.0	2280	1	PCT-US03-20322-211	Sequence 211, App
14	13	65.0	30	6	US-10-296-734-482	Sequence 482, App
15	13	65.0	2010	6	US-10-296-734-814	Sequence 814, App
16	10	50.0	25	6	US-10-685-435-25	Sequence 25, Appl
17	10	50.0	36	6	US-10-685-435-24	Sequence 24, Appl
18	10	50.0	44	6	US-10-685-435-23	Sequence 23, Appl
19	9	45.0	3033	6	US-10-009-002-5	Sequence 5, Appl
20	8	40.0	20	6	US-10-651-165-40	Sequence 40, Appl
21	7	35.0	30	6	US-10-296-734-478	Sequence 478, App
22	6	30.0	1547	6	US-10-679-063-15527	Sequence 15527, A
23	5	25.0	11	6	US-10-440-390-77	Sequence 77, Appl
24	5	25.0	11	6	US-10-440-390-82	Sequence 82, Appl
25	5	25.0	28	6	US-10-644-807-443	Sequence 443, App
26	5	25.0	36	5	US-09-350-841A-1084	Sequence 1084, Ap

27 5 25.0 36 5 US-09-350-369D-1703 Sequence 1703, Ap
28 5 25.0 36 5 US-09-350-369E-1703 Sequence 1703, Ap
29 5 25.0 36 6 US-10-351-641-1084 Sequence 1084, Ap
30 5 25.0 84 6 US-10-473-305-1293 Sequence 1293, Ap
31 5 25.0 126 6 US-10-425-114A-57646 Sequence 57646, A
32 5 25.0 157 1 PCT-US03-28227-48311 Sequence 4831, Ap
33 5 25.0 167 6 US-10-425-114A-58600 Sequence 58600, A
34 5 25.0 177 1 PCT-US03-32827-8 Sequence 8, Appl
35 5 25.0 177 6 US-10-685-288-14 Sequence 14, Appl
36 5 25.0 177 6 US-10-688-845-8 Sequence 8, Appl
37 5 25.0 191 6 US-10-425-114A-70769 Sequence 70769, A
38 5 25.0 202 4 US-08-657-749D-17 Sequence 17, Appl
39 5 25.0 202 7 US-60-478-196-3320 Sequence 3320, Ap
40 5 25.0 216 6 US-10-679-063-5587 Sequence 5587, Ap
41 5 25.0 216 5 US-09-979-932A-916 Sequence 916, App
42 5 25.0 225 6 US-10-425-114A-50995 Sequence 50995, A
43 5 25.0 236 5 US-09-979-932A-920 Sequence 920, App
44 5 25.0 237 6 US-10-425-114A-69808 Sequence 69808, A
45 5 25.0 238 6 US-10-679-063-6596 Sequence 6596, Ap

ALIGNMENTS

RESULT 1

US-10-651-165-39
; Sequence 39, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-39

Query Match 100.0% Score 20; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.3e-15; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 GNWFGCTWMNSTGFTKVCGA 20

RESULT 2

US-10-296-734-480
; Sequence 480, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2

```
; SEQ ID NO 480
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 37
US-10-296-734-480

Query Match      100.0%; Score 20; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.1e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GNWFGCTWMNSTGFTKVCGA 20
Db      9 GNWFGCTWMNSTGFTKVCGA 28

RESULT 3
PCT-US03-19834-3
; Sequence 3, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; TITLE OF INVENTION: METHODS FOR TREATING HYPERLIPIDEMIA
; FILE REFERENCE: IOWA:045W0
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-3

Query Match      100.0%; Score 20; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GNWFGCTWMNSTGFTKVCGA 20
Db      165 GNWFGCTWMNSTGFTKVCGA 184

RESULT 4
US-10-655-562-4
; Sequence 4, Application US/10655562
; GENERAL INFORMATION:
; APPLICANT: BISHOP-HURLEY, SHARON L.
; APPLICANT: SCHMIDT, FRANCIS J.
; APPLICANT: SMITH, ARNOLD L.
; TITLE OF INVENTION: PHAGE-DISPLAY PEPTIDES AS NOVEL ANTIMICROBIAL AGENTS
; TITLE OF INVENTION: AGAINST HAEMOPHILUS INFLUENZAE
; FILE REFERENCE: UVM0:022US
; CURRENT APPLICATION NUMBER: US/10/655,562
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: 60,409,909
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-655-562-4
```

```
Query Match      100.0%; Score 20; DB 6; Length 333;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GNWFGCTWMNSTGFTKVCGA 20
Db      165 GNWFGCTWMNSTGFTKVCGA 184

RESULT 5
PCT-US03-33610-4
; Sequence 4, Application PC/TUS0333610
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: SELBY, Mark
; APPLICANT: PALIARD, Xavier
; TITLE OF INVENTION: ACTIVATION OF HCV-SPECIFIC T-CELLS
; FILE REFERENCE: 2300-1612.60
; CURRENT APPLICATION NUMBER: PCT/US03/33610
; CURRENT FILING DATE: 2003-10-24
; PRIOR APPLICATION NUMBER: 10/281,341
; PRIOR FILING DATE: 2002-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
PCT-US03-33610-4

Query Match      100.0%; Score 20; DB 1; Length 637;
Best Local Similarity 100.0%; Pred. No. 3.4e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GNWFGCTWMNSTGFTKVCGA 20
Db      375 GNWFGCTWMNSTGFTKVCGA 394

RESULT 6
US-10-296-734-812
; Sequence 812, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 812
; LENGTH: 2011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette A
US-10-296-734-812

Query Match      100.0%; Score 20; DB 6; Length 2011;
Best Local Similarity 100.0%; Pred. No. 8.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GNWFGCTWMNSTGFTKVCGA 20
Db      400 GNWFGCTWMNSTGFTKVCGA 419
```



```
RESULT 7
PCT-US03-19834-2
; Sequence 2, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STABLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; FILE REFERENCE: IOWA:045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-2

Query Match      100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  GNWFGCTWMNSTGFTKVCGA 20
Db      547  GNWFGCTWMNSTGFTKVCGA 566

RESULT 8
US-10-296-734-406
; Sequence 406, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU P07761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a consensus polyprotein
US-10-296-734-406

Query Match      100.0%; Score 20; DB 6; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  GNWFGCTWMNSTGFTKVCGA 20
Db      547  GNWFGCTWMNSTGFTKVCGA 566

RESULT 9
US-10-296-734-810
; Sequence 810, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
```

```
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU P07761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 810
; LENGTH: 5985
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a savine
US-10-296-734-810

Query Match      100.0%; Score 20; DB 6; Length 5985;
Best Local Similarity 100.0%; Pred. No. 2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  GNWFGCTWMNSTGFTKVCGA 20
Db      399  GNWFGCTWMNSTGFTKVCGA 418

RESULT 10
US-10-664-391-9
; Sequence 9, Application US/10664391
; GENERAL INFORMATION:
; APPLICANT: Donnelly, John J.
; APPLICANT: Liu, Margaret A.
; APPLICANT: Shiver, John W.
; APPLICANT: Fu, Tong-Ming
; TITLE OF INVENTION: SYNTHETIC HEPATITIS C GENES
; FILE REFERENCE: 19732YPCA
; CURRENT APPLICATION NUMBER: US/10/664,391
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: PCT/US97/09884
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/033,534
; PRIOR FILING DATE: 1996-12-20
; PRIOR APPLICATION NUMBER: 60/020,494
; PRIOR FILING DATE: 1996-06-11
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-10-664-391-9

Query Match      80.0%; Score 16; DB 6; Length 347;
Best Local Similarity 100.0%; Pred. No. 4.1e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  GNWFGCTWMNSTGFTK 16
Db      165  GNWFGCTWMNSTGFTK 180

RESULT 11
US-10-664-391-11
; Sequence 11, Application US/10664391
; GENERAL INFORMATION:
; APPLICANT: Donnelly, John J.
; APPLICANT: Liu, Margaret A.
; APPLICANT: Shiver, John W.
; APPLICANT: Fu, Tong-Ming
; TITLE OF INVENTION: SYNTHETIC HEPATITIS C GENES
; FILE REFERENCE: 19732YPCA
; CURRENT APPLICATION NUMBER: US/10/664,391
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: PCT/US97/09884
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/033,534
; PRIOR FILING DATE: 1996-12-20
```

```
; PRIOR APPLICATION NUMBER: 60/020,494
; PRIOR FILING DATE: 1996-06-11
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-10-664-391-11

Query Match      80.0%; Score 16; DB 6; Length 539;
Best Local Similarity 100.0%; Pred. No. 5.8e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GNNFGCTWMNSTGFTK 16
        |||||
Db      357 GNNFGCTWMNSTGFTK 372

RESULT 12
PCT-US03-20409-3
; Sequence 3, Application PC/TUS0320409
; GENERAL INFORMATION:
; APPLICANT: Merck & Co., Inc.
; TITLE OF INVENTION: HEPATITIS C VIRUS PARTICLE FORMATION
; FILE REFERENCE: 21123 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/20409
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/393,167
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1026
; TYPE: PRT
; ORGANISM: HCV
PCT-US03-20409-3

Query Match      80.0%; Score 16; DB 1; Length 1026;
Best Local Similarity 100.0%; Pred. No. 9.6e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GNNFGCTWMNSTGFTK 16
        |||||
Db      547 GNNFGCTWMNSTGFTK 562

RESULT 13
PCT-US03-20322-211
; Sequence 211, Application PC/TUS0320322
; GENERAL INFORMATION:
; APPLICANT: Dana-Farber Cancer Institute, Inc., et al.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
; FILE REFERENCE: DFN-043PC
; CURRENT APPLICATION NUMBER: PCT/US03/20322
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 211
; LENGTH: 2280
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
PCT-US03-20322-211

Query Match      80.0%; Score 16; DB 1; Length 2280;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GNNFGCTWMNSTGFTK 16
        |||||

; PRIOR APPLICATION NUMBER: 60/020,494
; PRIOR FILING DATE: 1996-06-11
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-10-664-391-11

Query Match      80.0%; Score 16; DB 6; Length 539;
Best Local Similarity 100.0%; Pred. No. 5.8e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GNNFGCTWMNSTGFTK 16
        |||||
Db      357 GNNFGCTWMNSTGFTK 372

RESULT 14
US-10-296-734-482
; Sequence 482, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 482
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 38
US-10-296-734-482

Query Match      65.0%; Score 13; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 9.9e-08;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 WNNSTGFTKVCGA 20
        |||||
Db      1 WNNSTGFTKVCGA 13

RESULT 15
US-10-296-734-814
; Sequence 814, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 814
; LENGTH: 2010
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette B
US-10-296-734-814

Query Match      65.0%; Score 13; DB 6; Length 2010;
Best Local Similarity 100.0%; Pred. No. 2.7e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 WNNSTGFTKVCGA 20
        |||||
Db      391 WNNSTGFTKVCGA 403

Search completed: November 21, 2003, 22:12:57
Job time : 8.55 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:46:27 ; Search time 9.25 Seconds
(without alignments)
207.932 Million cell updates/sec

Title: US-09-973-025-83
Perfect score: 20
Sequence: 1 GNWFGCTWMNSTGFTKVCGA 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_76:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	3011	1 GNWVC3	genome polyprotein
2	20	100.0	3011	1 GNWVCH	genome polyprotein
3	20	100.0	3011	1 S40770	genome polyprotein
4	16	80.0	234	2 S32742	genome polyprotein
5	16	80.0	235	2 S32747	genome polyprotein
6	16	80.0	782	2 S19876	genome polyprotein
7	16	80.0	782	2 S18031	genome polyprotein
8	16	80.0	782	2 S19875	genome polyprotein
9	16	80.0	3010	1 GNWVTC	genome polyprotein
10	16	80.0	3010	1 GNWVCJ	genome polyprotein
11	16	80.0	3010	1 A45573	genome polyprotein
12	16	80.0	3010	1 S18030	genome polyprotein
13	16	80.0	3010	1 GNWVTV	genome polyprotein
14	14	70.0	350	2 S35631	genome polyprotein
15	14	70.0	782	2 S18032	genome polyprotein
16	14	70.0	3014	1 JC5620	genome polyprotein
17	11	55.0	640	2 JQ1584	genome polyprotein
18	11	55.0	3033	1 JQ1303	genome polyprotein
19	10	50.0	237	2 S32744	genome polyprotein
20	10	50.0	415	2 PC4407	envelope protein -
21	10	50.0	787	2 PQ0677	hypothetical prote
22	8	40.0	716	2 JQ1366	polyprotein - hepa
23	8	40.0	3033	1 GNWVJ8	genome polyprotein
24	6	30.0	146	2 F64449	hypothetical prote
25	6	30.0	148	2 S26450	hypothetical prote
26	6	30.0	168	2 H83075	type 4 fimbrial bi
27	6	30.0	212	2 S25996	hypothetical prote
28	6	30.0	386	2 A96532	hypothetical prote
29	6	30.0	554	2 C86366	protein F26P24.2 [

ALIGNMENTS

RESULT 1

GNWVC3

genome polyprotein - hepatitis C virus (strain HCV-1)
N:Contains: capsid protein C; envelope protein M; hepatitisvirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 19-Jan-2001
C:Accession: A39166; PQ0403; PQ0404
R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Co; Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A:Title: Genetic organization and diversity of the hepatitis C virus.
A:Reference number: A39166; MUID:91172826; PMID:1848704
A:Accession: A39166
A:Molecule type: mRNA
A:Residues: 1-3011 <CHO>
A:Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.J
J. Gen. Virol. 73, 1131-1141, 1992
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to
A:Reference number: PQ0393; MUID:92268871; PMID:1316939
A:Accession: PQ0403
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CHA>
A:Cross-references: DBJ:D10128
A:Experimental source: isolates E-b16
A:Accession: PQ0404
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CH2>
A:Experimental source: isolates E-b17
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5; status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEP>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitisvirin #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,536,576,623,645,1213,1255,2041,2077,22

Query Match 100.0% Score 20; DB 1; Length 3011;

Best Local Similarity 100.0%; Pred. No. 1.4e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20

|||||

Db 547 GNWFGCTWMNSTGFTKVCGA 566

RESULT 2

GNWVCH

genome polyprotein - hepatitis C virus (strain H)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

A:Note: host Homo sapiens (man)

C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001

R:Inchouape, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.

C:Accession: A36814; A41546

submitted to GenBank, July 1992

A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: C

A:Reference number: A36814

A:Accession: A36814

A:Molecule type: genomic RNA

A:Residues: 1-3011 <INC>

A:Cross-references: GB:M67463; NID:g329737; PIDN:AAA4534.1; PID:g329738

R:Inchouape, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.

Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991

A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: compar

A:Reference number: A41546; MUID:92052256; PMID:1658800

A:Contents: annotation

A:Note: neither amino acid nor nucleotide sequence is given

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura

F:1-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepacivirin #status predicted <NS3>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match 100.0%; Score 20; DB 1; Length 3011;

Best Local Similarity 100.0%; Pred. No. 1.4e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20

Db 547 GNWFGCTWMNSTGFTKVCGA 566

|||||

547 GNWFGCTWMNSTGFTKVCGA 566

RESULT 3

S40770

genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001

C:Accession: S40770; PC1285

submitted to the EMBL Data Library, March 1992

A:Reference number: S40770

A:Accession: S40770

A:Molecule type: genomic RNA

A:Residues: 1-3011 <OKA>

A:Cross-references: EMBL:D10749; NID:g221586; PIDN:BA01582.1; PID:g221587

R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,

Jpn. J. Exp. Med. 60, 167-177, 1990

A:Title: The 5'-terminal sequence of the hepatitis C virus genome.

A:Reference number: PC1284; MUID:91013116; PMID:2170712

A:Accession: PC1285

A:Molecule type: genomic RNA

A:Residues: 1-513 <OK2>

A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512

A:Experimental source: isolate HC-J1

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin

F:1-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepacivirin #status predicted <NS3>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match

Best Local Similarity

Matches 20; Conservative

0; Mismatches

0; Indels

0; Gaps

QY 1 GNWFGCTWMNSTGFTKVCGA 20

Db 547 GNWFGCTWMNSTGFTKVCGA 566

|||||

547 GNWFGCTWMNSTGFTKVCGA 566

RESULT 4

S32742

genome polyprotein - hepatitis C virus (isolate CR-1) (fragment)

N:Contains: envelope protein E2

C:Species: hepatitis C virus

A:Variety: isolate CR-1

C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S32742

R:Roggendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.

submitted to the EMBL Data Library, February 1993

A:Description: Variability of the envelope regions of HCV in European isolates and its

A:Reference number: S32741

A:Accession: S32742

A:Molecule type: genomic RNA

A:Residues: 1-234 <ROG>

A:Cross-references: EMBL:X72979; NID:g296102; PIDN:CAA51484.1; PID:g296103

A:Experimental source: isolate Cr-1

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: envelope protein; polyprotein

F:1-234/Product: envelope protein E2 #status predicted <MAT>

Query Match

Best Local Similarity

Matches 16; Conservative

0; Mismatches

0; Indels

0; Gaps

QY 1 GNWFGCTWMNSTGFTK 16

Db 178 GNWFGCTWMNSTGFTK 193

|||||

178 GNWFGCTWMNSTGFTK 193

RESULT 5

S32747

genome polyprotein - hepatitis C virus (isolate HU-1) (fragment)

N:Contains: envelope protein E2

C:Species: hepatitis C virus

A:Variety: isolate HU-1

C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S32747

R:Roggendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.

submitted to the EMBL Data Library, February 1993

A:Description: Variability of the envelope regions of HCV in European isolates and its

A:Reference number: S32741

A:Accession: S32747

A:Molecule type: genomic RNA

A:Residues: 1-235 <ROG>

A:Cross-references: EMBL:X72977; NID:g296112; PIDN:CAA51482.1; PID:g296113

A:Experimental source: isolate HU-1

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GNWFGCTWMNSTGFTK 16
    |||||
Db 547 GNWFGCTWMNSTGFTK 562

RESULT 10
GNWVCU
genome polyprotein - hepatitis C virus (strain J)
N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
C:Accession: A39253; PS0086
R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimoto
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients
A:Reference number: A39253; MUID:91088550; PMID:2175903
A:Accession: A39253
A:Molecule type: genomic RNA
A:Residues: 1-3010 <KAT>
A:Cross-references: GB:D90208; NID:g221610; PIDN:BAA14233.1; PID:g221611
R:Kato, N.; Ohkoshi, S.; Shimotohno, K.
Proc. Jpn. Acad. 65B, 219-223, 1989
A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari
A:Reference number: PS0085
A:Accession: PS0086
A:Molecule type: genomic RNA
A:Residues: 2650-2707 <KA2>
A:Experimental source: Japanese isolate
C:Comment: The cleavage sites of this polyprotein have not been determined.
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein NS2 #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,1213,1255,2041,2077,2240,2

Query Match 80.0%; Score 16; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 4.5e-10; Mismatches 0; Indels 0; Gaps 0;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GNWFGCTWMNSTGFTK 16
    |||||
Db 547 GNWFGCTWMNSTGFTK 562

RESULT 11
A45573
genome polyprotein - hepatitis C virus (strain JT)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: A45573
R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,
Virus Res. 23, 39-53, 1992
A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: s
A:Reference number: A45573; MUID:92295714; PMID:1318627
A:Accession: A45573
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3010 <TAN>

```

```

A:Cross-references: GB:D11168; GB:D01171; NID:g221612; PIDN:BAA01943.1; PID:g221613
A:Experimental source: HCV-JT
A>Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIP:106207)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein NS2 #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,1213,1255,2041,2077,2240,2

Query Match 80.0%; Score 16; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 4.5e-10; Mismatches 0; Indels 0; Gaps 0;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GNWFGCTWMNSTGFTK 16
    |||||
Db 547 GNWFGCTWMNSTGFTK 562

RESULT 12
S18030
genome polyprotein - hepatitis C virus (isolate JK1)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Variety: isolate JK1
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
C:Accession: S18030; S33570; A48332; S18029
R:Honda, M.; Kaneko, S.; Matsushita, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single patie
A:Reference number: S18028
A:Accession: S18030
A:Molecule type: genomic RNA
A:Residues: 1-3010 <HON>
A:Cross-references: EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59479
A:Experimental source: isolate JK1 from an individual
R:Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.
Arch. Virol. 128, 163-169, 1993
A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolated
A:Reference number: A48332; MUID:93119270; PMID:8380322
A:Accession: S33570
A:Molecule type: genomic RNA
A:Residues: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HOW>
A:Cross-references: EMBL:X61591
A>Note: this sequence is inconsistent with the nucleotide translation
A>Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320
as Trp, and TTC for residue 771 as Ser
A>Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBIP:121748)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein NS2 #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: DEXH motif
F:1316-1862/Product: nonstructural protein NS4a #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As

```

```

Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 WFGCTWMNSTGFTK 16
      |||
Db      168 WFGCTWMNSTGFTK 181

RESULT 15
S18032
genome polypeptide - hepatitis C virus (isolate JK4) (fragment)
N;Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein 3;
C;Species: hepatitis C virus
A;Variety: isolate JK4
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C;Accession: S18032
R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A;Description: Sequence analysis of putative structural regions of Hepatitis C
A;Reference number: S18029
A;Accession: S18032
A;Molecule type: genomic RNA
A;Residues: 1-782 <HON>
A;Cross-references: EMBL:X61594
A;Experimental source: isolate JK4
C;Superfamily: hepatitis C virus genome polypeptide
C;Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural protein 1; nonstructural protein 2; NS1/E2 protein 3;
F;1-191/Product: core protein #status predicted <MAT1>
F;192-383/Product: envelope protein 1 #status predicted <MAT2>
F;384-733/Product: NS1/E2 protein #status predicted <MAT3>
F;734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 70.0%; Score 14; DB 2; Length 782;
Best Local Similarity 100.0%; Pred. No. 2.9e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 WFGCTWMNSTGFTK 16
      |||
Db      549 WFGCTWMNSTGFTK 562

Search completed: November 21, 2003, 21:11:34
Job time : 10.25 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:52 ; Search time 4.85 Seconds
(without alignments)
193.925 Million cell updates/sec

Title: US-09-973-025-83

Perfect score: 20

Sequence: 1 GNWFGCTWMNSTGFTKVCGA 20

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	3011	1 POLG_HCV1	P26664 h genome po
2	20	100.0	3011	1 POLG_HCV1	P27958 h genome po
3	16	80.0	3010	1 POLG_HCVBK	P26663 h genome po
4	16	80.0	3010	1 POLG_HCVJA	P26662 h genome po
5	16	80.0	3010	1 POLG_HCVJT	Q00269 h genome po
6	16	80.0	3010	1 POLG_HCVTW	P29846 h genome po
7	14	70.0	737	1 POLG_HCVJS	P27960 hepatitis c
8	11	55.0	3033	1 POLG_HCVJ6	P26660 h genome po
9	8	40.0	737	1 POLG_HCVJ7	P27961 hepatitis c
10	8	40.0	3033	1 POLG_HCVJ8	P26661 h genome po
11	6	30.0	146	1 YB99_METJA	O58599 methanococ
12	6	30.0	148	1 YPZX_METTF	P29587 methanobact
13	6	30.0	556	1 SRW1_SCHPO	Q13286 schizosacch
14	6	30.0	605	1 CORO_CABEL	Q21624 caenorhabdi
15	6	30.0	662	1 ABFA_STRLI	P53627 streptomyc
16	5	25.0	22	1 PSP2_PHPYA	P80661 physcomitre
17	5	25.0	39	1 PSBJ_CVACA	Q9tm23 cyanidium c
18	5	25.0	65	1 Y00C_BPT4	Q01435 bacterioph
19	5	25.0	154	1 Y451_SYNY3	P74676 synecocyst
20	5	25.0	164	1 FLAV_HELPJ	Q92k53 helicobacte
21	5	25.0	177	1 IL7_HUMAN	P13322 homo sapien
22	5	25.0	181	1 APT_VIBCH	Q9kt52 vibrio chol
23	5	25.0	181	1 APT_VIBPA	Q87mq1 vibrio para
24	5	25.0	181	1 APT_VIBVU	Q8db25 vibrio vuln
25	5	25.0	211	1 YN08_CABEL	Q21984 caenorhabdi
26	5	25.0	220	1 NRG2 YEAST	P38082 saccharomyc
27	5	25.0	230	1 CHLM_SYNY3	Q55467 synecocyst
28	5	25.0	231	1 PSA2_CABEL	Q27488 caenorhabdi
29	5	25.0	238	1 T2RA_MOUSE	Q91ka3 mus musculu
30	5	25.0	252	1 TFXE_RHILT	P42727 rhizobium l
31	5	25.0	258	1 BDHA_ALCEU	Q9xeu2 alcaligenes
32	5	25.0	258	1 TPIS_STRCO	Q9z520 streptomyc
33	5	25.0	258	1 VSP1_TRIGA	O13059 trimeresuru

34	5	25.0	261	1 YAGS_STAAU	P55177 staphylococ
35	5	25.0	265	1 DAPB_AQUAE	O67061 aquifex aeo
36	5	25.0	288	1 DCMA_METSP	P21161 methylobact
37	5	25.0	289	1 TF_CAVPO	O91lu8 cavia porce
38	5	25.0	292	1 Y0RV_BACSU	O34416 bacillus su
39	5	25.0	307	1 UPK2_CLOAB	Q97kf6 clostridium
40	5	25.0	322	1 YMX7_CABEL	P34515 caenorhabdi
41	5	25.0	351	1 H0XN_ALCEU	P23516 alcaligenes
42	5	25.0	356	1 COA2_POVLY	P04011 lymphotropi
43	5	25.0	357	1 HPPD_PSESP	P80064 pseudomonas
44	5	25.0	366	1 TMA2_YEAST	Q04372 saccharomyc
45	5	25.0	371	1 MXIG_SHIFL	Q57332 shigella fl

ALIGNMENTS

RESULT 1

ID	POLG_HCV1	STANDARD;	PRT;	3011 AA.
AC	P26664;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Genome polyprotein (Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P87); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)).			
OS	Hepatitis C virus (isolate 1) (HCV).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.			
OX	NCBI_TaxID=11104;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91172826; PubMed=1848704;			
RA	Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;			
RT	"Genetic organization and diversity of the hepatitis C virus.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).			
CC	!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.			
CC	!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA) (N).			
CC	!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.			
CC	!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M62321; AAA45676.1; --
PIR; A39166; GNWVC3.
PDB; 1A4V; 16-FEB-99.
PDB; 1HEI; 25-NOV-98.
MEROPS; S29.001; --
MEROPS; S29.001; --
InterPro; IPR001410; DEAD.
InterPro; IPR002522; HCV_capsid.

```

DR InterPro: IPR002521; HCV core.
DR InterPro: IPR002519; HCV env.
DR InterPro: IPR002531; HCV NS1.
DR InterPro: IPR002518; HCV NS2.
DR InterPro: IPR004109; HCV NS3.
DR InterPro: IPR000745; HCV NS4.
DR InterPro: IPR001490; HCV NS4b.
DR InterPro: IPR002868; HCV NS5a.
DR InterPro: IPR002166; HCV RdRp.
DR InterPro: IPR001650; Helicase C.
DR InterPro: IPR007095; RNA pol DS PS.
DR InterPro: IPR007094; RNA pol pSvir.
DR Pfam: PF01543; HCV capsid; 1.
DR Pfam: PF01542; HCV core; 1.
DR Pfam: PF01539; HCV env; 1.
DR Pfam: PF01560; HCV NS1; 1.
DR Pfam: PF01538; HCV NS2; 1.
DR Pfam: PF02907; HCV NS3; 1.
DR Pfam: PF01006; HCV NS4a; 1.
DR Pfam: PF01001; HCV NS4b; 1.
DR Pfam: PF01506; HCV NS5a; 1.
DR Pfam: PF00271; Helicase C; 1.
DR Pfam: PF00998; Viral RdRp; 1.
DR ProDom: PD186062; HCV NS1; 1.
DR SMART: SM00487; DEXDC; 1.
KW Polyprotein: Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein: Glycoprotein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 131 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E1 (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 1007 1615 PROTEASE/HELICASE NS2 (POTENTIAL).
FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).
FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DESCH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 65F8C9447FCE5AF9 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.7e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GNNFGCTWMNSTGFTKVGCA 20

```

```

Db 547 GNNFGCTWMNSTGFTKVGCA 566
RESULT 2
POLG_HCVH STANDARD; PRT; 3011 AA.
AC P27958;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate H) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658800;
RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RT Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C
RT virus: comparison with American and Japanese isolates.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX MEDLINE=97331322; PubMed=9187654;
RA Yao N., Hesson T., Cabie M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain.";
RL Nat. Struct. Biol. 4:463-467(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE=98154321; PubMed=9493270;
RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RT Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding.";
RL Structure 6:89-100(1998).
CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

CC -----192-----5534.1; --
DR EMBL; M67463; AAA5534.1; --
DR PIR; A36814; GNVVCH.
DR PDB; 1HEI; 25-NOV-98.
DR PDB; 1A1V; 16-FEB-99.
DR PDB; 1A1R; 17-JUN-98.
DR MEROPS; S29.001; --
DR MEROPS; U39.001; --
DR TRANSFAC; T04155; --
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA pol DS PS.
DR InterPro; IPR007094; RNA pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; helicase C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
FT CHAIN 192 383 CAPSID PROTEIN C.
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
FT CHAIN 810 1026 PROTEIN P7.
FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5A.
FT TRANSMEM 347 369 NONSTRUCTURAL PROTEIN NS5B.
FT ACT SITE 1083 1083 POTENTIAL.
FT ACT SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT STRAND 1224 1226
FT TURN 1232 1233
FT TURN 1236 1238
FT TURN 1239 1246
FT TURN 1247 1248
FT STRAND 1251 1255
FT HELIX 1258 1271
FT TURN 1272 1272
FT STRAND 1277 1280
FT TURN 1281 1282
FT STRAND 1283 1285
FT STRAND 1291 1295
FT HELIX 1296 1301
FT TURN 1302 1303
FT STRAND 1312 1316
FT TURN 1317 1319
FT TURN 1323 1335
FT TURN 1336 1340
FT STRAND 1343 1347
FT TURN 1352 1353
FT TURN 1360 1361
FT STRAND 1362 1366
FT STRAND 1368 1368
FT STRAND 1373 1375
FT TURN 1376 1377
FT STRAND 1378 1380
FT HELIX 1382 1385
FT STRAND 1389 1393
FT HELIX 1397 1409
FT TURN 1410 1411
FT STRAND 1414 1417
FT TURN 1419 1420
FT STRAND 1432 1436
FT TURN 1438 1439
FT STRAND 1450 1453
FT STRAND 1456 1463
FT STRAND 1471 1478
FT STRAND 1480 1480
FT HELIX 1481 1488
FT TURN 1489 1490
FT STRAND 1497 1501
FT STRAND 1507 1507
FT STRAND 1511 1511
FT HELIX 1514 1527
FT HELIX 1532 1544
FT STRAND 1550 1550
FT HELIX 1555 1564
FT TURN 1570 1578
FT TURN 1579 1580
FT HELIX 1584 1597
FT TURN 1598 1598
FT HELIX 1606 1611
FT TURN 1614 1618
FT STRAND 1622 1623
FT STRAND 1627 1627
FT STRAND 1635 1636
FT HELIX 1640 1652
SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.7e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GNWFGCTWMNSTGFTKVCGA 20
Db 547 GNWFGCTWMNSTGFTKVCGA 566

RESULT 3
POLG_HCVBK STANDARD; PRT; 3010 AA.
ID POLG_HCVBK
AC P26663;
DT 01-AUG-1992 (Rel. 23, Created)

FT HELIX 1039 1047
 FT STRAND 1050 1050
 FT STRAND 1059 1063
 FT STRAND 1068 1074
 FT TURN 1075 1076
 FT STRAND 1077 1081
 FT HELIX 1082 1085
 FT TURN 1086 1087
 FT STRAND 1090 1092
 FT TURN 1093 1094
 FT STRAND 1095 1097
 FT STRAND 1101 1103
 FT TURN 1104 1107
 FT STRAND 1108 1112
 FT STRAND 1120 1120
 FT STRAND 1122 1122
 FT STRAND 1129 1133
 FT TURN 1135 1136
 FT STRAND 1139 1144
 FT STRAND 1149 1157
 FT HELIX 1158 1161
 FT TURN 1162 1163
 FT TURN 1165 1166
 FT STRAND 1168 1171
 FT TURN 1172 1174
 FT STRAND 1175 1186
 FT TURN 1187 1188
 FT STRAND 1189 1197
 FT HELIX 1198 1202
 FT TURN 1203 1204
 FT STRAND 1680 1698
 SQ SEQUENCE 3010 AA; 327189 MW; F8422D5BCCDFD9C CRC64;

Query Match 80.0%; Score 16; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred.No. 1.1e-10;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTK 16
 |||||
 DB 547 GNWFGCTWNNSTGFTK 562

RESULT 4
 POLG HCVJA STANDARD; PRT; 3010 AA.
 AC P26562;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP12) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin).
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Japanese) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91088550; PubMed=2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 RA Sugimura T., Shimotohno K.;
 RT "Molecular cloning of the human hepatitis C virus genome from
 RT Japanese patients with non-A, non-B hepatitis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RN [2]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=91192160; PubMed=1849488;
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K.,
 RA Ohkoshi S., Shimotohno K.;

"Molecular structure of the Japanese hepatitis C viral genome.";
 FEBS Lett. 280:325-328(1991).
 -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 (RNA) (N).
 -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA.
 -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; D90208; BAA14233.1; -;
 DR PIR; A39253; GNWVCJ.
 DR HSP; P26663; 1JXP.
 DR MEROPS; S29.001; -;
 DR MEROPS; U39.001; -;
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR00745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR007095; RNA pol_DS_Ps.
 DR InterPro; IPR007094; RNA pol_Psvir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107

FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;

Query Match 80.0%; Score 16; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 1.1e-10;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GNWFGCTWNSGTGFK 16
 |||||
 Db 547 GNWFGCTWNSGTGFK 562

RESULT 6
 ID POLG HCVTW STANDARD; PRT; 3010 AA.
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP58) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 DE Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_Taxid=31645;
 RN [1]
 RP MEDLINE=92230206; PubMed=1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 RT "The Taiwanese hepatitis C virus genome: sequence determination and
 RT mapping the 5' terminus of viral genomic and antigenomic RNA";
 RL Virology 188:102-113(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA) (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M84754; -; NOT_ANNOTATED_CDS.
 DR PIR; A40244; GNWVTW.
 DR PDB; 1N64; 25-FEB-03.
 DR PDB; 1NS3; 08-APR-98.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR007095; RNA pol DS PS.
 DR InterPro; IPR007094; RNA pol_PSVit.
 DR Pfam; PF01543; HCV_core; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure. 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 INIT_MET CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CORE PROTEIN (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT CHAIN 347 369 POTENTIAL.
 FT TRANSMEM 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 136 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 80.0%; Score 16; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 1.1e-10;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR007095; RNA pol^{BS}.
 DR InterPro; IPR007094; RNA pol^{PS}.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; helicase C; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT CHAIN 3034 369
 FT TRANSMEM
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2811 2811
 FT CARBOHYD 3033 3033
 SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BB9E CRC64;

Query Match 55.0%; Score 11; DB 1; Length 3033;
 Best Local Similarity 100.0%; Pred. No. 6.4e-05;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 WFGCTWNNSTG 13
 DB 551 WFGCTWNNSTG 561

RESULT 9
 POLG_HCVJ7 STANDARD; PRT; 737 AA.
 AC P27961;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
 DE protein (Envelope protein M); Major envelope protein E; Nonstructural
 DE proteins NS1 and NS2] (fragment).
 OS Hepatitis C virus (isolate HC-J7) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11114;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tauda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MENA.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D10077; BAA00971.1; -
 CC InterPro; IPR002522; HCV capsid.
 CC InterPro; IPR002521; HCV core.
 CC InterPro; IPR002519; HCV env.
 CC InterPro; IPR002531; HCV NS1.
 CC Pfam; PF01543; HCV capsid; 1.
 CC Pfam; PF01542; HCV core; 1.
 CC Pfam; PF01539; HCV env; 1.
 CC Pfam; PF01560; HCV NS1; 1.
 CC ProDom; PD186062; HCV NS1; 1.
 KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542

FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 737 737
 SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122F2 CRC64;
 Query Match 40.0%; Score 8; DB 1; Length 737;
 Best Local Similarity 100.0%; Pred. No. 0.056; Mismatches 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0;
 Qy 3 WFGCTWMN 10
 Db 551 WFGCTWMN 558
 RESULT 10
 POLG HCVJ8
 ID POLG HCVJ8 STANDARD; PRT; 3033 AA.
 AC P2661;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J8) (HCV).
 OC Viruses; sRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11115;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
 RA Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA}(N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D10988; BAA01761.1; -.
 CC PIR; A40250; GNWJ8.
 CC HSSP; P27958; 1HEI.
 CC MEROPS; S29.001; -.
 CC MEROPS; U39.001; -.
 CC InterPro; IPR001410; DEAD.
 CC InterPro; IPR002522; HCV capsid.
 CC InterPro; IPR002521; HCV_core.

DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRP.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXdc; 1.
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2359 2359
 FT CARBOHYD 2811 2811
 FT SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;
 Query Match 40.0%; Score 8; DB 1; Length 3033;
 Best Local Similarity 100.0%; Pred. No. 0.18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 3 WFGCTWMN 10
 Db 551 WFGCTWMN 558

```

RESULT 11
ID YB99 METJA STANDARD; PRT; 146 AA.
AC Q58599;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein MJ1199.
GN MJ1199.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Claydon R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- SIMILARITY: TO A.PERNIX APE2001.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; U67561; AAB99204.1; -.
DR PIR; F64449; F64449.
DR TIGR; MJ1199; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 146 AA; 16296 MW; 9934CB49198D85C CRC64;
Query Match 30.0%; Score 6; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 STGFTK 16
Db 108 STGFTK 113

RESULT 12
ID YPZX METTF STANDARD; PRT; 148 AA.
AC P29587;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE Hypothetical 16.7 kDa protein (ORF11).
OS Methanobacterium thermoformicicum.
OG Plasmid pF21.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 3720 / Z-245;
RX MEDLINE=93126090; PubMed=1336177;
RA Noelling J., van Eeden F.J.M., Eggen R.I.L., de Vos W.M.;
RT "Modular organization of related Archaeal plasmids encoding different
RT restriction-modification systems in Methanobacterium

```

```

thermoformicicum.";
RL Nucleic Acids Res. 20:6501-6507(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; X69367; CAA48440.1; -.
DR PIR; S30316; S26450.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 148 AA; 16658 MW; 77276336AA44853F CRC64;
Query Match 30.0%; Score 6; DB 1; Length 148;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 STGFTK 16
Db 52 STGFTK 57

RESULT 13
SRW1 SCHPO STANDARD; PRT; 556 AA.
ID SRW1 SCHPO
AC O13286;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE WD-repeat containing protein srw1 (Suppressor of rad/wee1).
GN SRW1 OR SFE9 OR SPAC144.13C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RX MEDLINE=98062330; PubMed=9398669;
RA Yanaguchi S., Murakami H., Okayama H.;
RT "A WD repeat protein controls the cell cycle and differentiation by
RT negatively regulating Cdc2/B-type cyclin complexes.";
RL Mol. Biol. Cell 8:2475-2486(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tullada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,

```

RA Shpakowski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC -!- FUNCTION: Has a role in cell differentiation and cell cycling by
 CC negatively regulating cig2 and cdc12-associated cdc2. Down-
 CC regulates the level of cdc13, particularly in a nitrogen deprived
 CC environment.
 CC -!- SIMILARITY: Contains 7 WD repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB005589; BAA22152.1; -;
 CC EMBL; AL132675; CAB59693.1; -;
 CC EMBL; T37680; T37680;
 CC GeneDB_Spombe; SPAC144.13c; -;
 CC InterPro; IPR001680; Fizzy.
 CC InterPro; IPR001680; WD40.
 CC Pfam; PF00400; WD40; 5.
 CC PRINTS; PR00320; GPROTEINRPT.
 CC ProDom; PD004563; Fizzy; 1.
 CC ProDom; PD000018; WD40; 1.
 CC SMART; SM00320; WD40; 5.
 CC PROSITE; PS00678; WD REPEATS 1; 1.
 CC PROSITE; PS0082; WD REPEATS 2; 4.
 CC PROSITE; PS0294; WD REPEATS_REGION; 1.
 CC Cell cycle; Repeat; WD repeat
 FT REPEAT 246 285 WD 1.
 FT REPEAT 289 328 WD 2.
 FT REPEAT 331 368 WD 3.
 FT REPEAT 372 411 WD 4.
 FT REPEAT 414 456 WD 5.
 FT REPEAT 458 499 WD 6.
 FT REPEAT 502 541 WD 7.
 SQ SEQUENCE 556 AA; 62060 MW; 9F9D13EC90F6964E CRC64;
 Query Match 30.0%; Score 6; DB 1; Length 556;
 Best Local Similarity 100.0%; Pred. No. 9;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 12 TGFTKV 17
 DB 88 TGFTKV 93
 RESULT 14
 CORO_CABEL STANDARD; PRT; 605 AA.
 AC Q21624;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Coronin-like protein.
 GN R01H0.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Lightning J.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: BINDS TO P-ACTIN (BY SIMILARITY).
 CC -!- SIMILARITY: Contains 3 WD repeats.
 CC -!- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z31590; CAA83461.1; -;
 CC EMBL; T37680; T37680;
 CC WormPep; R01H0.3; CE00590.
 CC InterPro; IPR001680; WD40.
 CC Pfam; PF00400; WD40; 3.
 CC SMART; SM00320; WD40; 3.
 CC PROSITE; PS00678; WD REPEATS 1; FALSE_NEG.
 CC PROSITE; PS0082; WD REPEATS_2; 3.
 CC PROSITE; PS0294; WD REPEATS_REGION; 1.
 CC KW Hypothetical protein; Actin-Binding; Repeat; WD repeat; Coiled coil.
 FT REPEAT 77 117 WD 1.
 FT REPEAT 127 167 WD 2.
 FT REPEAT 170 209 WD 3.
 FT DOMAIN 563 602 COILED COIL (POTENTIAL).
 SQ SEQUENCE 605 AA; 67060 MW; D98E1EC341507FB9 CRC64;
 Query Match 30.0%; Score 6; DB 1; Length 605;
 Best Local Similarity 100.0%; Pred. No. 9.6;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 STGFTK 16
 DB 231 STGFTK 236
 RESULT 15
 ABFA_STRLI STANDARD; PRT; 662 AA.
 AC PS3627;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Alpha-L-arabinofuranosidase (EC 3.2.1.55) (Arabinosidase) (ABF)
 DE (Alpha-L-AF).
 GN ABFA.
 OS Streptomyces lividans.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1916;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-23.
 RC STRAIN=66 / 1326;
 RX MEDLINE=94379974; PubMed=8092996;
 RA Manin C., Sharek F., Morosoli R., Kluepfel D.;
 RT "Purification and characterization of an alpha-L-arabinofuranosidase
 RT from Streptomyces lividans 66 and DNA sequence of the gene (abfa).";
 RL Biochem. J. 302:443-449(1994).
 CC -!- FUNCTION: Important role in the degradation of arabinose-
 CC containing hemicelluloses. Hydrolyzes rapidly the short-chain
 CC arabinoligosides from digestion of xylan with xylanases.
 CC Acts slowly on arabinan and arabinoxylan from wheat and rye flour
 CC to release L-arabinofuranose. Does not act on oat-spelts xylan or
 CC arabinogalactan.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing alpha-L-
 CC arabinofuranoside residues in alpha-L-arabinosides.
 CC -!- SUBUNIT: Oligomer.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- MISCELLANEOUS: Optimal enzyme activity occurs at 60 degrees
 CC Celsius and pH 6.0.
 CC -!- SIMILARITY: BELONGS TO FAMILY 51 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U04630; AAA61708.1; -.
DR PIR; S55274; S55274.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 662 AA; 72496 MW; DAAF66A577C1D6D1 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 662;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 TKVCGA 20
|||
Db 565 TKVCGA 570

Search completed: November 21, 2003, 20:59:59
Job time : 5.85 secs

THIS PAGE BLANK (11/15/70)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: November 21, 2003, 20:45:57 ; Search time 24.4 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-83
Perfect score: 20
Sequence: 1 GNWFGCTWMNSTGFTKVCGA 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL 23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_ivirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	20	100.0	137 12 Q914Q6	Q914Q6 hepatitis c
2	20	100.0	137 12 Q914Q5	Q914Q5 hepatitis c
3	20	100.0	137 12 Q914Q8	Q914Q8 hepatitis c
4	20	100.0	137 12 Q914Q4	Q914Q4 hepatitis c
5	20	100.0	137 12 Q914Q0	Q914Q0 hepatitis c
6	20	100.0	137 12 Q914Q3	Q914Q3 hepatitis c
7	20	100.0	137 12 Q914Q2	Q914Q2 hepatitis c
8	20	100.0	137 12 Q914Q1	Q914Q1 hepatitis c
9	20	100.0	137 12 Q914Q7	Q914Q7 hepatitis c
10	20	100.0	206 12 Q81571	Q81571 hepatitis c
11	20	100.0	420 12 Q98UN4	Q98UN4 hepatitis c
12	20	100.0	778 12 Q04185	Q04185 hepatitis c
13	20	100.0	778 12 Q04184	Q04184 hepatitis c
14	20	100.0	2436 12 Q81756	Q81756 hepatitis c
15	20	100.0	3011 12 Q91FE5	Q91FE5 hepatitis c
16	20	100.0	3011 12 Q36579	Q36579 hepatitis c

17	20	100.0	3011 12 Q36610	Q36610 hepatitis c
18	20	100.0	3011 12 Q03463	Q03463 hepatitis c
19	20	100.0	3011 12 Q9EL58	Q9EL58 hepatitis c
20	20	100.0	3011 12 Q36608	Q36608 hepatitis c
21	17	85.0	3011 12 Q36609	Q36609 hepatitis c
22	16	80.0	137 12 Q914T4	Q914T4 hepatitis c
23	16	80.0	137 12 Q914T3	Q914T3 hepatitis c
24	16	80.0	234 12 Q06620	Q06620 hepatitis c
25	16	80.0	235 12 Q06614	Q06614 hepatitis c
26	16	80.0	241 12 Q68448	Q68448 hepatitis c
27	16	80.0	241 12 Q68438	Q68438 hepatitis c
28	16	80.0	241 12 Q68434	Q68434 hepatitis c
29	16	80.0	241 12 Q68449	Q68449 hepatitis c
30	16	80.0	241 12 Q68435	Q68435 hepatitis c
31	16	80.0	241 12 Q68450	Q68450 hepatitis c
32	16	80.0	241 12 Q68453	Q68453 hepatitis c
33	16	80.0	241 12 Q68447	Q68447 hepatitis c
34	16	80.0	241 12 Q68454	Q68454 hepatitis c
35	16	80.0	376 12 Q81758	Q81758 hepatitis c
36	16	80.0	490 12 Q9PXN2	Q9PXN2 hepatitis c
37	16	80.0	490 12 Q9PXN0	Q9PXN0 hepatitis c
38	16	80.0	587 12 Q68518	Q68518 hepatitis c
39	16	80.0	681 12 Q68519	Q68519 hepatitis c
40	16	80.0	782 12 Q68953	Q68953 hepatitis c
41	16	80.0	782 12 Q68951	Q68951 hepatitis c
42	16	80.0	783 12 Q68950	Q68950 hepatitis c
43	16	80.0	802 12 Q68520	Q68520 hepatitis c
44	16	80.0	1008 12 Q89157	Q89157 hepatitis c
45	16	80.0	1008 12 Q89156	Q89156 hepatitis c

ALIGNMENTS

RESULT 1

Q914Q6 ID Q914Q6 PRELIMINARY; PRT; 137 AA.
AC Q914Q6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RC STRAIN=wt;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before and after liver transplantation."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431882; AAL30731.1; -;
DR InterPro; IPR002531; HCV NS1;
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 15042 MW; 08E5232119C91A09 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 8.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GNWFGCTWMNSTGFTKVCGA 20

Db 69 GNWFGCTWMNSTGFTKVCGA 88

RESULT 2

```

Q914Q5
ID Q914Q5 PRELIMINARY; PRT; 137 AA.
AC Q914Q5;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RC STRAIN=wt;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431883; AAL30732.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 15070 MW; 08F2132F29D01A09 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 8.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||
Db 69 GNWFGCTWMNSTGFTKVCGA 88

RESULT 3
Q914Q8
ID Q914Q8 PRELIMINARY; PRT; 137 AA.
AC Q914Q8;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RC STRAIN=wt;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431880; AAL30729.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 15070 MW; 08F2132F29D01A09 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 8.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||
Db 69 GNWFGCTWMNSTGFTKVCGA 88

Q914Q0
ID Q914Q0 PRELIMINARY; PRT; 137 AA.
AC Q914Q0;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RC STRAIN=wt;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431888; AAL30737.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 15078 MW; FF7E4F2219D33F35 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 8.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||
Db 69 GNWFGCTWMNSTGFTKVCGA 88

```

```

RESULT 4
Q914Q4
ID Q914Q4 PRELIMINARY; PRT; 137 AA.
AC Q914Q4;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RC STRAIN=wt;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431884; AAL30733.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 15070 MW; 08F2132F29D01A09 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 8.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||
Db 69 GNWFGCTWMNSTGFTKVCGA 88

RESULT 5
Q914Q0
ID Q914Q0 PRELIMINARY; PRT; 137 AA.
AC Q914Q0;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RC STRAIN=wt;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431888; AAL30737.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 15078 MW; FF7E4F2219D33F35 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 8.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWMNSTGFTKVCGA 20
|||
Db 69 GNWFGCTWMNSTGFTKVCGA 88

```


QY 1 GNWFGCTWNNSTGFTKVCGA 20
 |||||
 Db 69 GNWFGCTWNNSTGFTKVCGA 88

RESULT 6

Q914Q3 ID Q914Q3 PRELIMINARY; PRT; 137 AA.

AC Q914Q3; (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]_TaxID=11103;
 RP SEQUENCE FROM N.A.
 RC STRAIN=wt;
 RA Lyra A.C., Fan X., Di Bisceglie A.M.;
 RT "Genetic diversity of two putative cd81 binding regions of HCV before
 and after liver transplantation."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF431885; AAL30734.1; -.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 137
 SQ SEQUENCE 137 AA; 15074 MW; 188885A375CC1A09 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;

Best Local Similarity 100.0%; Pred. No. 8.9e-15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTKVCGA 20
 |||||
 Db 69 GNWFGCTWNNSTGFTKVCGA 88

RESULT 7

Q914Q2 ID Q914Q2 PRELIMINARY; PRT; 137 AA.

AC Q914Q2; (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]_TaxID=11103;
 RP SEQUENCE FROM N.A.
 RC STRAIN=wt;
 RA Lyra A.C., Fan X., Di Bisceglie A.M.;
 RT "Genetic diversity of two putative cd81 binding regions of HCV before
 and after liver transplantation."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF431886; AAL30735.1; -.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 137
 SQ SEQUENCE 137 AA; 15070 MW; 03A3EF7374DC0B08 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;

Best Local Similarity 100.0%; Pred. No. 8.9e-15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTKVCGA 20
 |||||
 Db 69 GNWFGCTWNNSTGFTKVCGA 88

RESULT 8

Q914Q1 ID Q914Q1 PRELIMINARY; PRT; 137 AA.

AC Q914Q1; (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]_TaxID=11103;
 RP SEQUENCE FROM N.A.
 RC STRAIN=wt;
 RA Lyra A.C., Fan X., Di Bisceglie A.M.;
 RT "Genetic diversity of two putative cd81 binding regions of HCV before
 and after liver transplantation."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF431887; AAL30736.1; -.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 137
 SQ SEQUENCE 137 AA; 15088 MW; 35ACEA66FB7731F2 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;

Best Local Similarity 100.0%; Pred. No. 8.9e-15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTKVCGA 20
 |||||
 Db 69 GNWFGCTWNNSTGFTKVCGA 88

RESULT 9

Q914Q7 ID Q914Q7 PRELIMINARY; PRT; 137 AA.

AC Q914Q7; (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]_TaxID=11103;
 RP SEQUENCE FROM N.A.
 RC STRAIN=wt;
 RA Lyra A.C., Fan X., Di Bisceglie A.M.;
 RT "Genetic diversity of two putative cd81 binding regions of HCV before
 and after liver transplantation."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF431888; AAL30730.1; -.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 137
 SQ SEQUENCE 137 AA; 15070 MW; 03A3EF7374DC0B08 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;

Best Local Similarity 100.0%; Pred. No. 8.9e-15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTKVCGA 20
 |||||
 Db 69 GNWFGCTWNNSTGFTKVCGA 88

RESULT 9

Q914Q7 ID Q914Q7 PRELIMINARY; PRT; 137 AA.

AC Q914Q7; (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]_TaxID=11103;
 RP SEQUENCE FROM N.A.
 RC STRAIN=wt;
 RA Lyra A.C., Fan X., Di Bisceglie A.M.;
 RT "Genetic diversity of two putative cd81 binding regions of HCV before
 and after liver transplantation."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF431889; AAL30731.1; -.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 137
 SQ SEQUENCE 137 AA; 15070 MW; 03A3EF7374DC0B08 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;

Best Local Similarity 100.0%; Pred. No. 8.9e-15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNWFGCTWNNSTGFTKVCGA 20
 |||||
 Db 69 GNWFGCTWNNSTGFTKVCGA 88

RESULT 9

Q914Q7 ID Q914Q7 PRELIMINARY; PRT; 137 AA.

AC Q914Q7; (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]_TaxID=11103;
 RP SEQUENCE FROM N.A.
 RC STRAIN=wt;
 RA Lyra A.C., Fan X., Di Bisceglie A.M.;
 RT "Genetic diversity of two putative cd81 binding regions of HCV before
 and after liver transplantation."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF431890; AAL30732.1; -.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 137
 SQ SEQUENCE 137 AA; 15070 MW; 03A3EF7374DC0B08 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;

Best Local Similarity 100.0%; Pred. No. 8.9e-15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DT	01-JUN-2001	(TrEMBLrel. 17, Created)	
DT	01-JUN-2001	(TrEMBLrel. 17, Last sequence update)	
DT	01-OCT-2002	(TrEMBLrel. 22, Last annotation update)	
DE	Genome polyprotein (Fragment).		
OS	Hepatitis C virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
OC	Hepacivirus.		
OC	NCBI_TaxID=11103;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Qia;		
RX	MEDLINE=20499063; PubMed=11044085;		
RA	Hadlock K.G., Lanford R.E., Perkins S., Rowe J., Yang Q., Levy S.,		
RA	Pileri P., Abrignani S., Fong S.K.;		
RT	"Human monoclonal antibodies that inhibit binding of hepatitis C virus		
RT	E2 protein to CD81 and recognize conserved conformational epitopes.";		
RN	J. Virol. 74:10407-10416(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Qia;		
RA	Keck Z.Y., Hadlock K.G., Yang Q.;		
RA	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AF348704; AAK32685.1; -		
DR	InterPro: IPR002519; HCV env.		
DR	InterPro: IPR002531; HCV NS1.		
DR	InterPro: IPR002518; HCV NS2.		
DR	Pfam: PF01539; HCV env; 1.		
DR	Pfam: PF01560; HCV NS1; 1.		
DR	Pfam: PF01538; HCV NS2; 1.		
DR	ProDom: PD186062; HCV NS1; 1.		
KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;		
KW	Polyprotein; Transmembrane.		
FT	NON TER 1 1		
FT	CHAIN 1 >37 E1.		
FT	CHAIN 0 38 >400 E2.		
FT	CHAIN 401 >420 P7.		
FT	NON TER 420 420		
SQ	SEQUENCE 420 AA; 45882 MW; F28DOC3BE0CD081E CRC64;		
Query Match	100.0%; Score 20; DB 12; Length 420;		
Best Local Similarity	100.0%; Pred. No. 2e-14; Mismatches 0; Gaps 0;		
Matches 20; Conservative 0; Indels 0;			
QY	1 GNWFGCTWNNSTGFTKVCGA 20		
DB	201 GNWFGCTWNNSTGFTKVCGA 220		
RESULT 12			
Q04185	PRELIMINARY; PRT; 778 AA.		
ID	Q04185; O81811;		
AC	Q04185; O81811;		
DT	01-NOV-1998 (TrEMBLrel. 08, Created)		
DT	01-NOV-1998 (TrEMBLrel. 08, Last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	Structural (Genome polyprotein) (Fragment).		
OS	Hepatitis C virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
OC	Hepacivirus.		
OC	NCBI_TaxID=11103;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=91195357; PubMed=1849654;		
RA	Ogata N., Alter H.J., Miller R.H., Purcell R.H.;		
RT	"Nucleotide sequence and mutation rate of the H strain of hepatitis C		
RT	virus.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 88:3392-3396(1991).		
CC	-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A		
CC	LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:		
CC	PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF		
CC	PROTEIN C AND MRNA (BY SIMILARITY).		
DR	EMBL; M62382; AAB02128.1; -		
DR	InterPro: IPR002522; HCV capsid.		

```
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR ProDom; PD186062; HCV NS2; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT CHAIN 514 >778 STRUCTURAL PROTEIN.
FT NON_TER 778
SQ SEQUENCE 778 AA; 85530 MW; 6E37E50AB97B6C32 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 778;
Best Local Similarity 100.0%; Pred. No. 3.1e-14; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

QY 1 GNWFGCTWNNSTGFTKVCQA 20
DB 547 GNWFGCTWNNSTGFTKVCQA 566

RESULT 13
Q04184
ID Q04184 PRELIMINARY; PRT; 778 AA.
AC Q04184; Q01810;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91195357; PubMed=1849654;
RA Ogata N., Alter H.J., Miller R.H., Purcell R.H.;
RT "Nucleotide sequence and mutation rate of the H strain of hepatitis C
virus.";
RL PROC. NATL. ACAD. SCI. U.S.A. 88:3392-3396(1991).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS.
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; M62381; AA02127.1; -.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR ProDom; PD186062; HCV NS2; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT CHAIN 514 >778 STRUCTURAL PROTEIN.
FT NON_TER 778
SQ SEQUENCE 778 AA; 85143 MW; F7709172CD003E39B CRC64;

Query Match 100.0%; Score 20; DB 12; Length 778;
Best Local Similarity 100.0%; Pred. No. 3.1e-14; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

QY 1 GNWFGCTWNNSTGFTKVCQA 20
DB 547 GNWFGCTWNNSTGFTKVCQA 566
```

```
RESULT 14
Q01756
ID Q01756 PRELIMINARY; PRT; 2436 AA.
AC Q01756;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX Choo Q.-L., Richman K., Han J.;
RT "The nucleotide sequence of the Hepatitis C viral genome.";
RL Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.
DR EMBL; M32084; AAA45677.1; -.
DR HSSP; P27958; IALV.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA pol_DS_Ps.
DR InterPro; IPR007094; RNA pol_PSVir.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
FT NON_TER 1
FT NON_TER 2436
SQ SEQUENCE 2436 AA; 264734 MW; D7B9872900BE3125 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 2436;
Best Local Similarity 100.0%; Pred. No. 7.2e-14; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

QY 1 GNWFGCTWNNSTGFTKVCQA 20
DB 97 GNWFGCTWNNSTGFTKVCQA 116

RESULT 15
Q01FE5
ID Q01FE5 PRELIMINARY; PRT; 3011 AA.
AC Q01FE5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
```

RP SEQUENCE FROM N.A.
RX MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
sequence.";
RL J. Gen. Virol. 82:1291-1297(2001).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS;
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
DR EMBL; AF271632; AAF81759.1; -.
DR HSSP; P27958; 1A1V.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_Psvir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR Pfam; PF0186062; HCV_NS1; 1.
DR PRODOM; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327124 MW; 2489CE74AC864E58 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 8.3e-14; Mismatches 0; Gaps 0;
Matches 20; Conservative 0; Indels 0;

QY 1 GNWFGCTWNSGTGFTKVCGA 20
|||
Db 547 GNWFGCTWNSGTGFTKVCGA 566

Search completed: November 21, 2003, 21:08:19
Job time : 24.45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 20:42:22 ; Search time 32.1 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-84

Perfect score: 20

Sequence: 1 GFTKVCAPPVCIGAGNNT 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A Geneseq_19Jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	AA194513	Hepatitis C virus
2	20	100.0	20	AA191007	HCV E2 peptide E2-
3	20	100.0	20	AA018713	Hepatitis C virus
4	10	50.0	26	AA1920752	Peptide 2 based on
5	10	50.0	30	AA084634	HCV HepC1a segment
6	10	50.0	30	AA084635	HCV HepC1a segment
7	10	50.0	96	AA077258	HCV bait polypepti
8	10	50.0	113	AA166632	Hepatitis C virus
9	10	50.0	113	AA01623	Protein encoded by

10	10	50.0	113	20	AAW30596	Hepatitis C virus
11	10	50.0	176	23	AB377257	HCV bait polypepti
12	10	50.0	179	21	AA18529	protein encoded by
13	10	50.0	192	19	AAW67009	HCV nucleocapsid c
14	10	50.0	250	23	AAU79220	Hepatitis C virus
15	10	50.0	254	22	AA568043	Amino acid sequenc
16	10	50.0	271	22	AA500445	HCV E2-634 HVR1 co
17	10	50.0	271	22	AA500446	HCV E2-634 HVR1-mu
18	10	50.0	278	14	AAW33997	Th E2/NS1 protein.
19	10	50.0	305	18	AAW00929	Recombinant HCV E2
20	10	50.0	333	14	AA140118	HGH-HCV-E2 fusion
21	10	50.0	337	16	AA179217	PHCV351-encoded AP
22	10	50.0	350	23	AA519891	Hepatitis C virus
23	10	50.0	353	14	AA18088	HCT27 E2/NS1 HV do
24	10	50.0	353	14	AA18088	HCT27 E2/NS1 HV do
25	10	50.0	363	22	AA568042	Amino acid sequenc
26	10	50.0	363	24	AA532880	Hepatitis C virus
27	10	50.0	363	24	ABF55567	Hepatitis C virus
28	10	50.0	367	14	AA133994	HCV E2/NS1 prote
29	10	50.0	367	14	AA140115	APP-HCV-E2 fusion
30	10	50.0	367	16	AA179218	PHCV167-encoded pr
31	10	50.0	377	16	AA179226	PHCV422-encoded AP
32	10	50.0	397	16	AA179220	PHCV419-encoded AP
33	10	50.0	399	14	AA140117	HGH-HCV-E2 fusion
34	10	50.0	402	14	AA144438	Sequence of glycop
35	10	50.0	402	14	AA144439	Sequence of glycop
36	10	50.0	402	14	AA144440	Sequence of glycop
37	10	50.0	403	13	AA130063	HCV E2/NS1 region.
38	10	50.0	403	20	AA122021	HCV E1 peptide seq
39	10	50.0	403	20	AA175482	Hepatitis C virus
40	10	50.0	409	14	AA133995	H77 E2/NS1 protein
41	10	50.0	409	14	AA133996	H90 E2/NS1 protein
42	10	50.0	410	16	AA179227	PHCV423-encoded AP
43	10	50.0	417	16	AA179228	PHCV424-encoded AP
44	10	50.0	434	16	AA179219	PHCV418-encoded AP
45	10	50.0	441	16	AA179230	PHCV429-encoded AP

ALIGNMENTS

RESULT 1

AA184513
ID AA184513 standard; peptide; 20 AA.

AC AA184513;

DT 06-JAN-1997 (first entry)

DE Hepatitis C virus peptide NS1-19 (residues 559-578).

DE Hepatitis C virus; HCV; immunogen; non-structural region; NS1;

KW immunodominant; T cell epitope; vaccine.

OS Hepatitis C virus.

PN WO9512677-A2.

PD 11-MAY-1995.

XX 28-OCT-1994; 94WO-EP03555.

XX 04-NOV-1993; 93EP-0402718.

XX (INNO-) INNOGENETICS NV.

XX Deleys R, Leroux-Roels G, Maertens G;

XX WPI; 1995-193822/25.

XX Hepatitis C Virus immunogenic polypeptide contg. a T-cell

PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in

PT production of vaccines, therapeutic agents, etc.

XX Example 4; Page 51; 105pp; English.

PS A series of overlapping peptides (including the present sequence) was

CC synthesised based on sequences in the core, E1 and E2/NS1 regions of

CC hepatitis C virus. The peptides were used as antigens in lympho-

CC proliferative assays to identify the main T-cell epitopes.

XX Sequence 20 AA;

SQ Query Match 100.0%; Score 20; DB 16; Length 20;

Best Local Similarity 100.0%; Pred. No. 7.5e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAGPPVCIGGAGNNT 20

Db 1 GFTKVCAGPPVCIGGAGNNT 20

RESULT 2

AA018713

ID AAR91007 standard; peptide; 20 AA.

XX AAR91007;

AC 25-SEP-1996 (first entry)

DT HCV E2 peptide E2-19B for competition studies.

DE HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;

XX serotype; reversed phase hybridisation assay; genotype; antigen; sera.

XX Synthetic.

OS WO9604385-A2.

XX 15-FEB-1996.

XX 31-JUL-1995; 95WO-EP03031.

XX 29-JUL-1994; 94EP-0870132.

XX (INNO-) INNOGENETICS NV.

XX Bosman F, Buyse M, De Martynoff G, Maertens G;

XX WPI; 1996-129401/13.

XX Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope

XX proteins - in presence of disulphide bond cleavage agent, to

XX produce proteins suitable for direct use in vaccines or diagnostic

XX assays of HCV

XX Example 7; Page 67; 146pp; English.

PS AAR90965-P90980 and AAR90995-P91015 represent synthetic hepatitis C

XX virus (HCV) E1 and E2 peptides used in competition studies. This

XX sequence represents a synthetic E2 peptide, and corresponds to residues

XX 559-578 of the E2 protein sequence. These sequences are useful for in

XX vitro monitoring of HCV disease, or prognosis of the response to

XX interferon treatment of patients suffering from HCV infection. These

XX sequences compete with the proteins produced by AAT12704-T12709 and

XX AAT12961-T12974, which are included in vectors for the production of

XX recombinant E1, E2, and E1/E2. The recombinant proteins can then be

XX isolated and purified by carrying out a disulphide bond cleavage, or a

XX reduction step with a disulphide bond cleavage agent, after lysis of

XX recombinant host cells. The constructs containing the purified HCV

XX envelope proteins can be used for vaccinating humans against HCV, for in

XX vitro detection of HCV antibodies in a sample, and in a serotyping assay

XX for detecting one or more serological types of HCV present in a

XX biological sample. The constructs can also be immobilised on a solid

XX substrate and incorporated into a reversed phase hybridisation assay for

XX determining the presence or the genotype of HCV. The new-purification

XX method preserves the conformation of the recombinantly expressed E1, E2

XX and E1/E2, and eliminates contaminating proteins. Antigens isolated

XX using this method are more reactive with human sera than those isolated

XX by known techniques.

XX Sequence 20 AA;

SQ Query Match 100.0%; Score 20; DB 17; Length 20;

Best Local Similarity 100.0%; Pred. No. 7.5e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAGPPVCIGGAGNNT 20

Db 1 GFTKVCAGPPVCIGGAGNNT 20

RESULT 3

AA018713

ID AAO18713 standard; Peptide; 20 AA.

XX AAO18713;

AC 24-OCT-2002 (first entry)

DT Hepatitis C virus E2 protein derived peptide E2-19B.

DE Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide;

XX immunostimulant; vaccine.

XX Hepatitis C virus.

OS WO200255548-A2.

XX 18-JUL-2002.

XX 11-JAN-2002; 2002WO-EP00219.

XX 11-JAN-2001; 2001US-260699P.

XX 30-AUG-2001; 2001US-315768P.

XX (INNO-) INNOGENETICS NV.

XX Maertens G, Bosman F, Buyse M;

XX WPI; 2002-599657/64.

XX New therapeutic vaccine compositions comprising at least one purified

XX recombinant hepatitis C virus (HCV) single or specific oligomeric

XX recombinant envelope protein E1 or E2, useful for immunizing humans

XX from HCV infection

XX Example 7; Page 229; 243pp; English.

PS The present invention relates to new therapeutic vaccine compositions for

XX inducing hepatitis C virus (HCV)-specific antibodies, comprising a

XX composition containing at least one purified recombinant HCV single or

XX specific oligomeric recombinant envelope proteins selected from an E1 and

XX an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are

XX useful for inducing HCV-specific antibodies or for immunising humans

XX against HCV. The recombinant HCV E1 and/or E2 proteins are useful as

XX vaccines or therapeutics, in HCV screening and confirmatory antibody

XX tests, for raising antibodies, in the preparation of medicament, and for

XX in vitro monitoring of HCV disease or prognosing the response to

XX treatment of patients suffering from HCV infection. The present sequence

XX is a peptide derived from the proteins of the invention.

XX Sequence 20 AA;

SQ Query Match 100.0%; Score 20; DB 23; Length 20;

Best Local Similarity 100.0%; Pred. No. 7.5e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAGPPVCIGGAGNNT 20

Db 1 GFTKVCAPPVCIAGAGNT 20

RESULT 4
AAR20752 AAR20752 standard; Protein; 26 AA.
ID AC AAR20752;
XX AC
XX AC
XX 25-MAR-2003 (updated)
DT 05-MAY-1992 (first entry)
XX
XX Peptide 2 based on immunoreactive region of Hepatitis C virus.
DE
XX
XX Non-A, non-B hepatitis virus; non-structural protein; vaccine.
KW
XX
XX Synthetic.
OS
XX EP468527-A.
PN
XX
XX 29-JAN-1992.
PD
XX
XX 26-JUL-1991; 91EP-0112620.
PF
XX
XX 24-JUN-1991; 91US-0719819.
PR
XX 26-JUL-1990; 90US-058799.
PR
XX 07-FEB-1991; 91US-0651735.
PR
XX 11-MAR-1991; 91US-0667275.
XX
XX (UNBI-) UNITED BIOMEDICAL INC.
PA
XX
XX Chang YW, Hosein B;
PI
XX
XX WPI; 1992-034279/05.
DR
XX
XX New synthetic peptide specific for HCV antibodies - for detection
PT of HCV or NANBHV e.g. by enzyme-linked immunosorbent assay and is
PT immunogen for preparation of vaccines
XX
XX Claim 1; Page 88; 98pp; English.
PS
XX This peptide is one of 19 specifically claimed antigens based on the
CC immunoreactive regions of the envelope protein and non-structural
CC proteins NS-1, NS-2, NS-3 and NS-5 for the Hepatitis C virus. The
CC invention also concerns analogues, segments, mixtures, conjugates
CC and polymers of these peptides. The C-terminal amino acid may be
CC amidated. See AAR20751-R20782.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
XX
SQ Sequence 26 AA;

Query Match 50.0%; Score 10; DB 13; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
Db 16 GFTKVCAPP 25

RESULT 5
AAU84634
ID AAU84634 standard; Peptide; 30 AA.
XX
XX AC AAU84634;
XX
XX 08-MAY-2002 (first entry)
DT
XX HCV HepC1a segment 37.
DE
XX
KW Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
KW viral infection; human immunodeficiency virus; melanoma;

KW bacterial infection; Salmonella; Legionella; parasitic infection;
KW Trypanosoma; Toxoplasma; Giardia.
XX
XX Hepatitis C virus.
OS
XX WO200190197-A1.
PN
XX 29-NOV-2001.
PD
XX
XX 25-MAY-2001; 2001WO-AU00622.
PF
XX
XX 26-MAY-2000; 2000AU-0007761.
PR
XX
XX (AUS) UNIV AUSTRALIAN NAT.
PA
XX
XX Thomson SA, Ramshaw IA;
PI
XX
XX WPI; 2002-147575/19.
DR
XX N-PSDB; ABK36472.

New synthetic polypeptides having several different segments of at least one parent polypeptide linked together differently compared to the linkage in the parent polypeptide, for inducing immune response against a pathogen or cancer -

Example 2; Fig 26; 364pp; English.

The invention relates to a new synthetic polypeptide (I) comprising several different segments of at least one parent polypeptide linked together in a different relationship relative to their linkage in the parent polypeptide to impede, abrogate or otherwise alter at least one function associated with the parent polypeptide and for inducing an immune response against a pathogen or cancer. Also included are a synthetic polynucleotide encoding and a computer system for designing the synthetic polypeptides. The synthetic polypeptides and polynucleotides are referred to as a Savine. The synthetic polypeptide is useful for modulating immune responses preferably directed against a pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix, colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone liver, oesophagus, brain, testicle, uterus), as potentiating agents. Compositions comprising the polypeptide may be used in the treatment or prophylaxis against viral (such as infections caused by HIV (human immunodeficiency virus), hepatitis, influenza, Japanese encephalitis virus, Epstein-Barr virus and respiratory syncytial virus), bacterial (e.g., infections caused by Neisseria, Meningococcal, Haemophilus, CC Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic (e.g., infections caused by Plasmodium, Schistosoma, Leishmania, Trypanosoma, Toxoplasma and Giardia) infections. The present sequence is a peptide derived from a parent protein used to construct a savine of the invention.

SQ Sequence 30 AA;

Query Match 50.0%; Score 10; DB 23; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
Db 21 GFTKVCAPP 30

RESULT 6
AAU84635
ID AAU84635 standard; Peptide; 30 AA.
XX
XX AC AAU84635;
XX
XX 08-MAY-2002 (first entry)
DT
XX HCV HepC1a segment 38.
DE
XX
KW Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;

KW viral infection; human immunodeficiency virus; melanoma;
 KW bacterial infection; Salmonella; Legionella; parasitic infection;
 KW Trypanosoma; Toxoplasma; Giardia.

XX Hepatitis C virus.

XX WO200190197-A1.

XX 29-NOV-2001.

XX 25-MAY-2001; 2001WO-AU00622.

XX 26-MAY-2000; 2000AU-0007761.

XX (AUS) UNIV AUSTRALIAN NAT.

XX Thomson SA, Ramshaw IA;

XX WPI; 2002-147575/19.

XX N-PSDB; ABK36473.

XX New synthetic polypeptides having several different segments of at
 PT least one parent polypeptide linked together differently compared to
 PT the linkage in the parent polypeptide, for inducing immune response
 PT against a pathogen or cancer -

XX Example 2; Fig 26; 364pp; English.

XX The invention relates to a new synthetic polypeptide (I) comprising
 CC several different segments of at least one parent polypeptide linked
 CC together in a different relationship relative to their linkage in the
 CC parent polypeptide to impede, abrogate or otherwise alter at least one
 CC function associated with the parent polypeptide and for inducing an
 CC immune response against a pathogen or cancer. Also included are a
 CC synthetic polynucleotide encoding and a computer system for
 CC designing the synthetic polypeptides. The synthetic polypeptide is
 CC useful for modulating immune responses preferably directed against a
 CC pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix,
 CC colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone
 CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.
 CC Compositions comprising the polypeptide may be used in the treatment or
 CC prophylaxis against viral (such as infections caused by HIV (human
 CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
 CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
 CC (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,
 CC Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic
 CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
 CC Trypanosoma, Toxoplasma and Giardia) infections. The present
 CC sequence is a peptide derived from a parent protein used to
 CC construct a vaccine of the invention.

XX Sequence 30 AA;

Query Match 50.0%; Score 10; DB 23; Length 30;
 Best Local Similarity 100.0%; Pred. No. 0.0058;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPP 10
 |||||
 Db 6 GFTKVCGAPP 15

RESULT 7

ABB77258
 ID ABB77258 standard; Protein; 96 AA.

XX ABB77258;

XX 28-JUN-2002 (first entry)

XX HCV bait polypeptide 6.

XX

KW SID; selected interacting domain; HCV; hepatitis C virus;
 KW liver disease; liver cancer; virucide; hepatotropic; antiinflammatory;
 KW antibacterial.

XX Hepatitis C virus strain H77.

XX EP1178116-A1.

XX 06-FEB-2002.

XX 03-AUG-2000; 2000EP-0402225.

XX 03-AUG-2000; 2000EP-0402225.

XX (HYBR-) HYBRIGENICS SA.

XX Legrain P, Whiteside S, Wojcik J;

XX WPI; 2002-208115/27.

XX N-PSDB; ABL55590.

XX New selected interacting domain polypeptides and polynucleotides,
 PT useful for treating or preventing infections or pathologies caused by
 PT hepatitis C virus (HCV) or those linked to HCV infection -

XX Claim 26; SEQ ID 82; 61pp + sequence listing; English.

XX The invention relates to nucleic acids encoding polypeptides which are
 CC termed SID polypeptides (selected interacting domain). These polypeptides
 CC are the final products of a double selection method involving a first
 CC step of selection of Hepatitis C virus (HCV)-derived polynucleotides
 CC through a two-hybrid system, and a second selection step involving an
 CC alignment between the different polynucleotides selected at the first
 CC step. The activity of polypeptides of the invention may be described as,
 CC virucide, hepatotropic, antiinflammatory and antibacterial. The
 CC polypeptide, polynucleotide and compositions comprising them are useful
 CC for treating or preventing viral or a bacterial infection, specifically
 CC infections or pathologies caused by HCV, or those pathologies linked to
 CC HCV infection. These may include liver disease and liver cancer. The
 CC current sequence represents a HCV bait polypeptide.
 CC Note: The sequence data for this patent is not represented in the
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.

XX Sequence 96 AA;

Query Match 50.0%; Score 10; DB 23; Length 96;
 Best Local Similarity 100.0%; Pred. No. 0.015;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPP 10
 |||||
 Db 73 GFTKVCGAPP 82

RESULT 8

AAR66632
 ID AAR66632 standard; Protein; 113 AA.

XX AAR66632;

XX 25-MAR-2003 (updated)

XX 31-AUG-1995 (first entry)

XX Hepatitis C virus J1 NS1 domain consensus protein.

XX Hepatitis C virus J1 NS1 domain; anti-HCV vaccine development;
 KW non-A non-B virus; diagnostic polypeptides; HCV probes.

XX Hepatitis C virus.

XX US5372928-A.

XX


```

PD 13-DEC-1994.
XX PF 24-FEB-1994; 94US-0201066.
XX PR 15-SEP-1989; 89US-0408045.
XX PR 21-DEC-1989; 89US-0456142.
XX PR 04-JAN-1991; 91US-0637380.
XX PR 02-AUG-1993; 93US-0101280.
XX PR 24-FEB-1994; 94US-0201066.
XX (CHIR ) CHIRON CORP.
PA (NAHE-) NAT INST OF HEALTH JAPAN.
XX Cha T, Han J, Houghton M, Irvine BD, Kolberg JA;
PI Miyamura T, Saito I, Weiner AJ;
XX N-PSDB; AAQ79775.
DR WPI; 1995-030306/04.
DR N-PSDB; AAQ79775.
XX Method of detecting hepatitis C virus polynucleotide - utilises
PT probe based on DNA of new HCV isolates J1 and J7
XX Claim 1; Fig 17; 45pp; English.
XX AAQ79775 encodes AAR66632 the prod. of the hepatitis C virus (HCV)
CC J1 NS1 domain consensus sequence. They can be used to provide
CC new oligonucleotides and polypeptides for use in diagnostics,
CC recombinant protein prodn. and anti-HCV vaccine development.
CC (Updated on 25-MAR-2003 to correct PF field.)
XX
XX SQ Sequence 113 AA;
Query Match 50.0%; Score 10; DB 16; Length 113;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GFTKVCGAPP 10
Db 14 GFTKVCGAPP 23

RESULT 9
ID AAY01623 standard; Protein; 113 AA.
AC AAY01623;
XX 21-JUN-1999 (first entry)
DE Protein encoded by the coding strand of the HCV J1 NS1 domain.
KW HCV; J7 isolate; J1 isolate; HCV1; immunoassay; asiatic strain;
KW diagnosis; HCV infection; blood screening; immunisation;
XX antiviral.
XX Hepatitis C virus.
XX US5871903-A.
XX 16-FEB-1999.
XX 08-MAY-1995; 95US-0436965.
XX 04-JAN-1991; 91US-0637380.
XX 15-SEP-1989; 89US-0408045.
XX 21-DEC-1989; 89US-0456142.
XX 02-AUG-1993; 93US-0101280.
XX 24-FEB-1994; 94US-0201066.
XX 03-NOV-1994; 94US-0334255.
XX 08-MAY-1995; 95US-0436965.
XX (CHIR ) CHIRON CORP.
PA (NAHE-) NAT INST OF HEALTH JAPAN.

```

```

XX Miyamura T, Saito I;
PI WPI; 1999-166619/14.
DR N-PSDB; AAX26742.
XX Immunoassays for Asiatic strains of hepatitis C virus - for
PT diagnosis of infection and screening blood supplies
XX Disclosure; Fig 17; 43pp; English.
XX
XX The present sequence is encoded by the consensus sequence of the coding
CC strand of a new hepatitis C virus (HCV), J1, NS1 domain. The J1 and J7
CC (also a new HCV isolate) isolates comprise sequences which are distinct
CC from the prototype HCV isolates, HCV1. The specification describes
CC immunoassays for HCV based on antigens from Asiatic strains not
CC cross-reactive with HCV-1. The assays are used for diagnosis of HCV
CC infection and to screen donated blood. The anti-HCV antibodies are also
CC useful therapeutically and prophylactically (passive immunisation); in
CC screening for antiviral agents; for isolation, purification and
CC identification of non-A, non-B hepatitis virus (e.g. by affinity
CC chromatography) and to raise anti-idiotypic antibodies (useful for
CC treatment or diagnosis and to determine immunogenic regions of the
CC HCV antigens).
XX
XX SQ Sequence 113 AA;
Query Match 50.0%; Score 10; DB 20; Length 113;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GFTKVCGAPP 10
Db 14 GFTKVCGAPP 23

RESULT 10
AAW30596
ID AAW30596 standard; Protein; 113 AA.
XX AC AAW30596;
XX 10-MAY-1999 (first entry)
DE Hepatitis C virus J1 isolate NS1 region.
KW HCV; non-A, non-B hepatitis; NANBH; diagnosis; vaccine; antigen;
KW antibody; immunoassay; assay; non-structural domain; NS1.
XX Hepatitis C virus.
XX US5856437-A.
XX 05-JAN-1999.
XX 03-NOV-1994; 94US-0334255.
XX 04-JAN-1991; 91US-0637380.
XX 15-SEP-1989; 89US-0408045.
XX 21-DEC-1989; 89US-0456142.
XX 02-AUG-1993; 93US-0101280.
XX 24-FEB-1994; 94US-0201066.
XX 03-NOV-1994; 94US-0334255.
XX (CHIR ) CHIRON CORP.
PA (NAHE-) NAT INST OF HEALTH JAPAN.
XX Cha T, Han J, Houghton M, Irvine BD, Kolberg JA;
PI Miyamura T, Saito I, Weiner AJ;
XX WPI; 1999-105191/09.
DR N-PSDB; AAX00446.
XX

```

PT Antigenic polypeptides from J1 and J7 hepatitis C virus isolates -
PT useful as immunoassay reagents, for raising antibodies and as
PT vaccine components
XX
PS Example 5; Fig 17; 44pp; English.
XX
CC This polypeptide comprises a non-structural domain 1 (NS1) region
CC of novel Japanese isolate J1 of hepatitis C virus (HCV), as deduced
CC from an amplified DNA clone (see AAK00446). The sequence shows homology
CC to prototype HCV-1. The invention provides new Japanese isolates,
CC J1 and J7, of HCV. The new isolates have nucleotide and amino acid
CC sequences which are distinct from the prototype HCV-1 isolate.
CC These differences can be exploited for use in diagnostics for
CC NANBH, recombinant protein production and vaccine development.
CC Claimed antigenic polypeptides (see AAW30583-87) can be used: (i) as
CC immunoassay reagents, or standards, to detect HCV antibodies, e.g.
CC for diagnosing infection or screening donated blood; (ii) to
CC generate specific antibodies (used for detecting the corresponding
CC polypeptide, to screen for antiviral agents, for virus isolation
CC and for passive immunisation); (iii) in protective or therapeutic
CC vaccines, and (iv) for isolation of non-A, non-B viruses.
XX
SQ Sequence 113 AA;
Query Match 50.0%; Score 10; DB 20; Length 113;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GFTKVCGAPP 10
Db 14 GFTKVCGAPP 23
|||||
RESULT 11
ABB77257
ID ABB77257 standard; Protein; 176 AA.
AC ABB77257;
DT 28-JUN-2002 (first entry)
DE HCV bait polypeptide 5.
XX
XX SID; selected interacting domain; HCV; hepatitis C virus;
KW liver disease; liver cancer; virucide; hepatotropic; antiinflammatory;
KW antibacterial.
XX
XX Hepatitis C virus strain H77.
OS
XX
PN EPI178116-A1.
XX
PD 06-FEB-2002.
XX
XX 03-AUG-2000; 2000EP-0402225.
PF
XX
XX 03-AUG-2000; 2000EP-0402225.
PR
XX
XX (HYBR-) HYBRIGENICS SA.
PA
XX
PI Legrain P, Whiteside S, Wojcik J;
XX
XX WPI; 2002-208115/27.
DR
XX
XX N-PSDB; ABL55589.
XX
XX New selected interacting domain polypeptides and polynucleotides,
PT useful for treating or preventing infections or pathologies caused by
PT hepatitis C virus (HCV) or those linked to HCV infection -
XX
XX Claim 26; SEQ ID 81; 61pp + sequence listing; English.
XX
XX The invention relates to nucleic acids encoding polypeptides which are
CC termed SID polypeptides (selected interacting domain). These polypeptides
CC are the final products of a double selection method involving a first

CC step of selection of Hepatitis C virus (HCV)-derived polynucleotides
CC through a two-hybrid system, and a second selection step involving an
CC alignment between the different polynucleotides selected at the first
CC step. The activity of polypeptides of the invention may be described as,
CC virucide, hepatotropic, antiinflammatory and antibacterial. The
CC polypeptide, polynucleotide and compositions comprising them are useful
CC for treating or preventing viral or a bacterial infection, specifically
CC infections or pathologies caused by HCV, or those pathologies linked to
CC HCV infection. These may include liver disease and liver cancer. The
CC current sequence represents a HCV bait polypeptide.
CC Note: The sequence data for this patent is not represented in the
CC specification, but is based on sequence information supplied by the
CC European Patent Office.
XX
SQ Sequence 176 AA;
Query Match 50.0%; Score 10; DB 23; Length 176;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GFTKVCGAPP 10
Db 91 GFTKVCGAPP 100
|||||
RESULT 12
AAB18529
ID AAB18529 standard; Protein; 179 AA.
XX
AC AAB18529;
XX
DT 15-JAN-2001 (first entry)
DE
XX Protein encoded by a novel hepatitis C virus cDNA clone 131.
XX Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
KW viral infectivity; viral replication.
XX
XX Hepatitis C virus.
OS
XX
XX EPI034785-A2.
PN
XX
PD 13-SEP-2000.
XX
XX 16-MAR-1990; 2000EP-0109602.
PF
XX
XX 17-MAR-1989; 89US-0325338.
PR
XX 20-APR-1989; 89US-0341334.
PR
XX 18-MAY-1989; 89US-0355002.
PR
XX 16-MAR-1990; 90EP-0302866.
XX
XX (CHIR) CHIRON CORP.
PA
XX
XX Houghton M, Choo Q, Kuo G;
PI
XX
XX WPI; 2000-566891/53.
DR
XX N-PSDB; AAA75285.
XX
XX Novel composition comprising a hepatitis C virus antisense
PT polynucleotide which is complementary to or corresponds to a sense
PT strand of the virus genome, and selectively hybridises to it -
XX
XX Example; Fig 5; 75pp; English.
XX
XX The specification describes a pharmaceutical composition which
CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
CC HCV is characterized by a positive stranded RNA genome which has
CC 40% homology at the polypeptide level to a HCV polypeptide. The
CC antisense polynucleotide binds to cellular polynucleotides which
CC enhance and/or are required for viral infectivity, replicative
CC ability or chronicity. The antisense polynucleotides may also be
CC designed to bind with high specificity, to be of increased stability,
CC to be stable and to have low toxicity. The composition also comprises

CC an agent which causes viral RNA to be inactive. The composition
 CC is used for preventing HCV replication in a system. The present
 CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
 CC course of the invention.

XX SQ Sequence 179 AA;

Query Match 50.0%; Score 10; DB 21; Length 179;
 Best Local Similarity 100.0%; Pred. No. 0.025;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
 |||||
 Db 49 GFTKVCAPP 58

RESULT 13

AAW67009
 ID AAW67009 standard; protein; 192 AA.

XX AC AAW67009;

XX DT 02-MAR-1999 (first entry)

XX DE HCV nucleocapsid core protein.

XX KW Hepatitis C virus; HCV; nucleocapsid; core protein; analogue; antibody;
 KW non-structural protein; thioamide bond; peptide bond.

XX OS Hepatitis C virus.

XX FH Key Location/Qualifiers
 FT Misc-difference 79

FT /note= "given in specification as Try"

XX JPI0226698-A.

XX PD 25-AUG-1998.

XX PF 19-FEB-1997; 97JP-0034702.

XX PR 19-FEB-1997; 97JP-0034702.

XX PA (KYOW) KYOWA MEDEX KK.

XX DR WPI; 1998-515103/44.

XX PT Determination of antibody in sample - uses peptide analog absorbed
 PT or chemically bound on carrier as antigen

XX PS Disclosure; Page 4; 13pp; Japanese.

XX CC This sequence represents the Hepatitis C virus (HCV) nucleocapsid core
 CC protein. The invention relates to peptide analogues derived from HCV
 CC proteins, e.g. AAW67417-W67426, which can be used for the determination
 CC of anti-HCV antibodies in a sample. Preferably the peptide analogues
 CC contain one or more thioamide peptide bonds where at least one oxygen
 CC atom of the peptide bond is replaced by sulphur atom. The peptide
 CC analogues can be adsorbed or chemically bound to a carrier.

XX SQ Sequence 192 AA;

Query Match 50.0%; Score 10; DB 19; Length 192;
 Best Local Similarity 100.0%; Pred. No. 0.027;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
 |||||
 Db 109 GFTKVCAPP 118

RESULT 14

AAU79220

ID AAU79220 standard; Protein; 250 AA.

XX AC AAU79220;

XX DT 15-JUL-2002 (first entry)

XX DE Hepatitis C Virus (HCV) delta-delta E2 genotype 1a protein.

XX KW Hepatitis C Virus E2; HCV E2; virucide; hepatotropic; IL-2;
 KW antiinflammatory; HCV infection; interleukin-2; gamma-interferon;
 KW granulocyte macrophage-colony stimulating factor; GM-CSF;
 KW delta-delta E2 genotype 1a.

XX OS Hepatitis C Virus.

XX PN WO200222155-A1.

XX PD 21-MAR-2002.

XX PF 13-SEP-2001; 2001WO-US28767.

XX PR 13-SEP-2000; 2000US-230927P.

XX PA (HAWA-) HAWAII BIOTECHNOLOGY GROUP INC.

XX PI Nakano ET, Clements DE, Humphreys T;

XX DR WPI; 2002-383102/41.

XX DR N-PSDB; ABK49386.

XX FT New immunogenic polypeptide comprising hepatitis C virus E2 polypeptide
 FT useful for treating hepatitis C virus infection and for providing
 FT immune protection against virus infection

XX PS Claim 4; Fig 6; 84pp; English.

XX CC The invention relates to a secreted polypeptide comprising hepatitis C
 CC virus (HCV) E2 polypeptide lacking all or a portion of its membrane
 CC spanning domain so that the E2 polypeptide is capable of secretion into
 CC growth medium when expressed recombinantly in a host cell. The
 CC polypeptide may also lack a portion of its C-terminus. The HCV E2
 CC secreted polypeptide is useful for producing anti-HCV antibodies. A
 CC purified immunogenic polypeptide comprising HCV E2 is useful for treating
 CC HCV infection and for providing immune protection against HCV infection
 CC by administering it to a subject having or at risk of having HCV
 CC infection or in need of protection. The method further comprises
 CC administering an immunomodulatory agent such as interleukin-2 (IL-2),
 CC granulocyte macrophage-colony stimulating factor (GM-CSF) or
 CC gamma-interferon. The polypeptide is useful as a vaccine, and with other
 CC HCV proteins to form a multi-component HCV vaccine for prophylactic or
 CC therapeutic treatment of HCV infection. This sequence represents an N-
 CC and C-terminally truncated HCV delta-delta E2 genotype 1a protein.

XX SQ Sequence 250 AA;

Query Match 50.0%; Score 10; DB 23; Length 250;
 Best Local Similarity 100.0%; Pred. No. 0.033;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
 |||||
 Db 148 GFTKVCAPP 157

RESULT 15

AAB68043

ID AAB68043 standard; protein; 254 AA.

XX AC AAB68043;

XX DT 29-JUN-2001 (first entry)

XX DE Amino acid sequence of water soluble variant of envelope E2 protein.

XX Eo protein; HCV; envelope protein; E2 protein; HCV infection;
KW HCV attachment.
XX
XX Synthetic.
OS Hepatitis C virus.
XX
PN WO200122984-A1.
XX
XX 05-APR-2001.
XX
XX 26-SEP-2000; 2000WO-US26395.
XX
XX 29-SEP-1999; 99US-0407430.
XX
XX (UYCO) UNIV COLUMBIA NEW YORK.
XX
XX Worman HJ, Mamiya N;
XX
XX WPI; 2001-273486/28.
XX
XX Treating or preventing hepatitis C virus infection in a subject,
PT involves administering hepatitis C virus envelope protein E2 binding
PT agents
XX
XX Claim 5; Fig 8; 46pp; English.
XX
XX The present sequence represents a water soluble variant of a Hepatitis C
CC virus (HCV) envelope E2 protein. Agents that bind to the HCV E2 protein
CC (such as the human Eo protein), and so inhibit the attachment of HCV onto
CC cells (especially liver cells), are used to treat HCV infections in
CC mammals, in particular humans. The specification also describes a method
CC for identifying a compound which can be used for treating or preventing
CC HCV in a subject and which can inhibit the attachment of HCV onto cells
CC by inhibiting the binding of HCV envelope E2 protein to a cellular
CC protein associated with HCV attachment and entry into cells. The method
CC comprises incubating the compound, HCV envelope E2 protein or its variant
CC and a cellular protein capable of specifically binding to the HCV E2
CC protein under suitable reaction conditions; determining the interactions
CC between HCV envelope E2 protein and cellular protein in the presence and
CC absence of the compound; and comparing the interaction to identify a
CC compound which can inhibit the attachment of HCV onto cells.
XX
SQ Sequence 254 AA;

Query Match 50.0%; Score 10; DB 22; Length 254;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPP 10
Db 154 GFTKVCGAPP 163

Search completed: November 21, 2003, 20:58:05
Job time : 32.15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.6 Seconds
(without alignments)
79.832 Million cell updates/sec

Title: US-09-973-025-84

Perfect score: 20

Sequence: 1 GFTKVCAPPVIGGAGNNT 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-08-612-973-84
2	20	100.0	20	3	US-08-927-597-84
3	20	100.0	20	4	US-08-635-886C-40
4	10	50.0	26	1	US-08-262-037-2
5	10	50.0	34	1	US-08-262-037-41
6	10	50.0	42	1	US-08-262-037-42
7	10	50.0	50	1	US-08-262-037-43
8	10	50.0	88	1	US-08-440-103-27
9	10	50.0	88	1	US-08-440-542-27
10	10	50.0	88	1	US-08-231-368-27
11	10	50.0	88	1	US-08-440-210-27
12	10	50.0	88	4	US-09-046-604-27
13	10	50.0	179	3	US-08-444-818-77
14	10	50.0	278	1	US-08-440-103-15
15	10	50.0	278	1	US-08-440-542-15
16	10	50.0	278	1	US-08-231-368-15
17	10	50.0	278	1	US-08-440-210-15
18	10	50.0	278	4	US-09-046-604-15
19	10	50.0	305	3	US-08-478-073-2
20	10	50.0	333	1	US-08-453-552-12
21	10	50.0	333	2	US-08-710-637-12
22	10	50.0	333	5	PCT-US93-00907-12
23	10	50.0	337	1	US-08-188-281B-7
24	10	50.0	337	5	PCT-US94-07280-7
25	10	50.0	337	5	PCT-US95-01087-7
26	10	50.0	353	1	US-08-440-103-31
27	10	50.0	353	1	US-08-440-103-32

28	10	50.0	353	1	US-08-440-542-31	Sequence 31, Appl
29	10	50.0	353	1	US-08-440-542-32	Sequence 32, Appl
30	10	50.0	353	1	US-08-231-368-31	Sequence 31, Appl
31	10	50.0	353	1	US-08-231-368-32	Sequence 32, Appl
32	10	50.0	353	1	US-08-440-210-31	Sequence 31, Appl
33	10	50.0	353	1	US-08-440-210-32	Sequence 32, Appl
34	10	50.0	353	4	US-09-046-604-31	Sequence 31, Appl
35	10	50.0	353	4	US-09-046-604-32	Sequence 32, Appl
36	10	50.0	367	1	US-08-188-281B-9	Sequence 9, Appl
37	10	50.0	367	1	US-08-453-552-6	Sequence 6, Appl
38	10	50.0	367	1	US-08-440-103-17	Sequence 17, Appl
39	10	50.0	367	1	US-08-440-542-17	Sequence 17, Appl
40	10	50.0	367	1	US-08-231-368-17	Sequence 17, Appl
41	10	50.0	367	1	US-08-440-210-17	Sequence 17, Appl
42	10	50.0	367	2	US-08-710-637-6	Sequence 6, Appl
43	10	50.0	367	4	US-09-046-604-17	Sequence 17, Appl
44	10	50.0	367	5	PCT-US93-00907-6	Sequence 6, Appl
45	10	50.0	367	5	PCT-US94-07280-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1 0
US-08-612-973-84
; Sequence 84, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-612-973-84

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFTKVCAPPVIGGAGNNT 20
|||||

Db 1 GFTKVCAPPVCIGGAGNNT 20

RESULT 2

US-08-927-597-84
; Sequence 84, Application US/08927597
; Patent No. 6245503
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P. C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927.597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612.973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32.205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-927-597-84

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPPVCIGGAGNNT 20

Db 1 GFTKVCAPPVCIGGAGNNT 20

RESULT 3

US-08-635-886C-40
; Sequence 40, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635.886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555

; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40
; LENGTH: 20
; TYPE: PRP
; ORGANISM: hepatitis C virus
US-08-635-886C-40

Query Match 100.0%; Score 20; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPPVCIGGAGNNT 20

Db 1 GFTKVCAPPVCIGGAGNNT 20

RESULT 4

US-08-262-037-2
; Sequence 2, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262.037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-262-037-2

```
Query Match      50.0%; Score 10; DB 1; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
Db 16 GFTKVCAPP 25

RESULT 5
US-08-262-037-41
; Sequence 41, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
; US-08-262-037-41

Query Match      50.0%; Score 10; DB 1; Length 34;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
Db 24 GFTKVCAPP 33

RESULT 6
US-08-262-037-42
; Sequence 42, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
; US-08-262-037-42

Query Match      50.0%; Score 10; DB 1; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.0028;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
Db 32 GFTKVCAPP 41

RESULT 7
US-08-262-037-43
; Sequence 43, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
```

CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/262,037
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/719,819
FILING DATE: 24-June-1991
APPLICATION NUMBER: 07/667,275
FILING DATE: 11-Mar-1991
APPLICATION NUMBER: 07/651,735
FILING DATE: 07-Feb-1991
APPLICATION NUMBER: 07/558,799
FILING DATE: 26-July-1990
APPLICATION NUMBER: 07/510,153
FILING DATE: 16-April-1990
ATTORNEY/AGENT INFORMATION:
NAME: Maria C. H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4043 US3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 amino acids
TYPE: Amino acid
STRANDEDNESS:
TOPOLOGY: Unknown
US-08-262-037-43

Query Match 50.0%; Score 10; DB 1; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.0033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPP 10
Db 40 GFTKVCGAPP 49

RESULT 8
US-08-440-103-27
Sequence 27, Application US/08440103
Patent No. 5670152
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,103
FILING DATE: 12-MAY-1995

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-103-27
Query Match 50.0%; Score 10; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPP 10
Db 6 GFTKVCGAPP 15

RESULT 9
US-08-440-542-27
Sequence 27, Application US/08440542
Patent No. 5670153
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,542
FILING DATE: 12-MAY-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 amino acids
TYPE: amino acid
STRANDEDNESS: single


```
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-542-27

Query Match      50.0%; Score 10; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GFTKVCGAPP 10
Db      6 GFTKVCGAPP 15

RESULT 10
US-08-231-368-27
; Sequence 27, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,210
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-210-27

Query Match      50.0%; Score 10; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GFTKVCGAPP 10
Db      6 GFTKVCGAPP 15

RESULT 12
US-09-046-604-27
; Sequence 27, Application US/09046604
; Patent No. 6303292
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/046,604
; FILING DATE:
; CLASSIFICATION:
```

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/231,368
;; FILING DATE: 14-MAR-1995
;; APPLICATION NUMBER: US 07/759,575
;; FILING DATE: 13-SEP-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: McClung, Barbara G.
;; REGISTRATION NUMBER: 33,113
;; REFERENCE/DOCKET NUMBER: 0205.001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (510) 601-2708
;; TELEFAX: (510) 655-3542
;; INFORMATION FOR SEQ ID NO: 27:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 88 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-09-046-604-27

Query Match 50.0%; Score 10; DB 4; Length 88;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPP 10
Db 6 GFTKVCGAPP 15

RESULT 13
US-08-444-818-77
; Sequence 77, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 179 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-444-818-77

Query Match 50.0%; Score 10; DB 3; Length 179;
Best Local Similarity 100.0%; Pred. No. 0.0096;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 GFTKVCGAPP 10
Db 49 GFTKVCGAPP 58

RESULT 14
US-08-440-103-15
; Sequence 15, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,103
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-103-15

Query Match 50.0%; Score 10; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPP 10
Db 190 GFTKVCGAPP 199

RESULT 15
US-08-440-542-15
; Sequence 15, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Chiron Corporation
;; STREET: 4560 Horton Street
;; CITY: Emeryville
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94608
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/440,542
;; FILING DATE: 12-MAY-1995
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/231,368
;; FILING DATE:
;; APPLICATION NUMBER: US 07/759,575
;; FILING DATE: 13-SEP-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: McClung, Barbara G.
;; REGISTRATION NUMBER: 33,113
;; REFERENCE/DOCKET NUMBER: 0205.001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (510) 601-2708
;; TELEFAX: (510) 655-3542
;; INFORMATION FOR SEQ ID NO: 15:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 278 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-440-542-15

Query Match 50.0%; Score 10; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFTKVCAPP 10
Db 190 GFTKVCAPP 199

Search completed: November 21, 2003, 21:15:16
Job time : 10.6 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time 19.5 Seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-84

Perfect score: 20

Sequence: 1 GFTKVCAPPVCGAGNNT 20

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/FCI_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/FCIUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09D_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09E_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	10	US-09-973-025-84
2	20	100.0	20	11	US-09-899-303-84
3	20	100.0	20	11	US-09-995-808-84
4	20	100.0	20	11	US-09-995-860-84
5	20	100.0	20	12	US-09-995-791-84
6	10	50.0	96	10	US-09-921-397-82
7	10	50.0	176	10	US-09-921-397-81
8	10	50.0	250	10	US-09-952-572-8
9	10	50.0	254	10	US-09-407-430-3
10	10	50.0	350	10	US-09-929-955-4
11	10	50.0	350	14	US-10-104-966-4
12	10	50.0	363	10	US-09-407-430-2
13	10	50.0	363	12	US-10-128-587A-97
14	10	50.0	363	15	US-10-128-590-97
15	10	50.0	637	12	US-10-187-257-4

16	10	50.0	637	12	US-10-265-083-2	Sequence 2, Appli
17	10	50.0	2894	10	US-09-941-611-23	Sequence 23, Appl
18	10	50.0	2894	15	US-10-044-995-23	Sequence 23, Appl
19	10	50.0	3011	9	US-09-742-659-4	Sequence 4, Appli
20	10	50.0	3011	9	US-09-916-359-2	Sequence 2, Appli
21	10	50.0	3011	10	US-09-238-076-20	Sequence 20, Appl
22	10	50.0	3011	10	US-09-952-572-9	Sequence 9, Appli
23	10	50.0	3011	10	US-09-929-955-1	Sequence 1, Appli
24	10	50.0	3011	10	US-09-747-419-20	Sequence 20, Appl
25	10	50.0	3011	11	US-09-891-894-3	Sequence 3, Appli
26	10	50.0	3011	11	US-09-995-937-20	Sequence 20, Appl
27	10	50.0	3011	11	US-09-917-563-20	Sequence 20, Appl
28	10	50.0	3011	12	US-10-184-150-3	Sequence 3, Appli
29	10	50.0	3011	14	US-10-104-966-1	Sequence 1, Appli
30	10	50.0	3011	15	US-10-259-275-20	Sequence 20, Appl
31	10	50.0	3011	16	US-10-232-643-6	Sequence 2, Appli
32	610	59.0	3012	10	US-09-238-076-2	Sequence 2, Appli
33	10	50.0	3012	11	US-09-995-937-2	Sequence 2, Appli
34	10	50.0	3012	11	US-09-917-563-2	Sequence 2, Appli
35	8	40.0	19	10	US-09-973-025-85	Sequence 85, Appl
36	8	40.0	19	11	US-09-899-303-85	Sequence 85, Appl
37	8	40.0	19	11	US-09-995-808-85	Sequence 85, Appl
38	8	40.0	19	11	US-09-995-860-85	Sequence 85, Appl
39	8	40.0	19	12	US-09-995-791-85	Sequence 85, Appl
40	8	40.0	20	10	US-09-973-025-83	Sequence 83, Appl
41	8	40.0	20	11	US-09-899-303-83	Sequence 83, Appl
42	8	40.0	20	11	US-09-995-808-83	Sequence 83, Appl
43	8	40.0	20	11	US-09-995-860-83	Sequence 83, Appl
44	8	40.0	20	12	US-09-995-791-83	Sequence 83, Appl
45	8	40.0	30	16	US-10-318-200-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1

US-09-973-025-84
; Sequence 84, Application US/09973025
; Publication NO. US20020182706A1

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

BOSMAN, FONS

DE MARTYNOFF, GUY

BUYSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/973,025

FILING DATE: 10-Oct-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-973-025-84

Query Match 100.0%; Score 20; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFTKVCAPPVCIGGAGNNT 20
Db 1 GFTKVCAPPVCIGGAGNNT 20

RESULT 2
US-09-899-303-84
; Sequence 84, Application US/09899303
; Publication No. US20030036110A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-899-303-84

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFTKVCAPPVCIGGAGNNT 20
|||||

Db 1 GFTKVCAPPVCIGGAGNNT 20

RESULT 3
US-09-995-808-84
; Sequence 84, Application US/09995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 84
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-84

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFTKVCAPPVCIGGAGNNT 20
|||||

Db 1 GFTKVCAPPVCIGGAGNNT 20

RESULT 4
US-09-995-860-84
; Sequence 84, Application US/09995860
; Publication No. US20030118603A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and therapeutic use.
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 84
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-84

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFTKVCAPPVCIGGAGNNT 20
|||||

Db 1 GFTKVCAPPVCIGGAGNNT 20

RESULT 5
US-09-995-791-84
; Sequence 84, Application US/09995791
; Publication No. US20030147918A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and therapeutic use.
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 84

```
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-84

Query Match          100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPPVCIGGAGNNT 20
    |||||
Db 1 GFTKVCAPPVCIGGAGNNT 20
    |||||

RESULT 6
US-09-921-397-82
; Sequence 82, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 82
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-82

Query Match          50.0%; Score 10; DB 10; Length 96;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
    |||||
Db 73 GFTKVCAPP 82
    |||||

RESULT 7
US-09-921-397-81
; Sequence 81, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 81
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-81

Query Match          50.0%; Score 10; DB 10; Length 176;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
    |||||

; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-84

Query Match          100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPPVCIGGAGNNT 20
    |||||
Db 1 GFTKVCAPPVCIGGAGNNT 20
    |||||

RESULT 8
US-09-952-572-8
; Sequence 8, Application US/09952572
; Patent No. US20020119495A1
; GENERAL INFORMATION:
; APPLICANT: HAWAII BIOTECHNOLOGY GROUP, Inc.
; APPLICANT: NAKANO, Eileen
; APPLICANT: CLEMENTS, David
; APPLICANT: HUMPHREYS, Tom
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITION OF HEPATITIS C
; FILE REFERENCE: HAWBIO1100
; CURRENT APPLICATION NUMBER: US/09/952,572
; CURRENT FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: US 60/230,927
; PRIOR FILING DATE: 2000-09-13
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-09-952-572-8

Query Match          50.0%; Score 10; DB 10; Length 250;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
    |||||
Db 148 GFTKVCAPP 157
    |||||

RESULT 9
US-09-407-430-3
; Sequence 3, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Norman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; TITLE OF INVENTION: HEPATITIS C VIRUS INFECTION
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-407-430-3

Query Match          50.0%; Score 10; DB 10; Length 254;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
    |||||
Db 154 GFTKVCAPP 163
    |||||

RESULT 10
US-09-929-955-4
; Sequence 4, Application US/09929955
; Patent No. US20020136740A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
```

```
; FILE REFERENCE: TRIPEP.23AUS2
; CURRENT APPLICATION NUMBER: US/09/929,955
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/225,767
; PRIOR FILING DATE: 2000-08-17
; SOFTWARE: FastSeq for Windows Version 4.0
; NUMBER OF SEQ ID NOS: 49
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-09-929-955-4

Query Match      50.0%; Score 10; DB 10; Length 350;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GFTKVCGAPP 10
      |||||
DB      180 GFTKVCGAPP 189

RESULT 11
US-10-104-966-4
; Sequence 4, Application US/10104966
; Publication No. US20020155124A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.23AUSC1
; CURRENT APPLICATION NUMBER: US/10/104,966
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-10-104-966-4

Query Match      50.0%; Score 10; DB 14; Length 350;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GFTKVCGAPP 10
      |||||
DB      180 GFTKVCGAPP 189

RESULT 12
US-09-407-430-2
; Sequence 2, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; TITLE OF INVENTION: HEPATITIS C VIRUS INFECTION
; FILE REFERENCE: 0575/54805
```

```
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-09-407-430-2

Query Match      50.0%; Score 10; DB 10; Length 363;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GFTKVCGAPP 10
      |||||
DB      176 GFTKVCGAPP 185

RESULT 13
US-10-128-587A-97
; Sequence 97, Application US/10128587A
; Publication No. US20030152940A1
; GENERAL INFORMATION:
; APPLICANT: Imogenetics N.V.
; TITLE OF INVENTION: Constructs and methods for expression of recombinant HCV envelope
; TITLE OF INVENTION: proteins
; FILE REFERENCE: 134 PCT
; CURRENT APPLICATION NUMBER: US/10/128,587A
; CURRENT FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: hepatitis C virus
US-10-128-587A-97

Query Match      50.0%; Score 10; DB 12; Length 363;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GFTKVCGAPP 10
      |||||
DB      176 GFTKVCGAPP 185

RESULT 14
US-10-128-590-97
; Sequence 97, Application US/10128590
; Publication No. US20030108561A1
; GENERAL INFORMATION:
; APPLICANT: Imogenetics N.V.
; TITLE OF INVENTION: Core-glycosylated HCV envelope proteins
; FILE REFERENCE: 135 PCT
; CURRENT APPLICATION NUMBER: US/10/128,590
; CURRENT FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-128-590-97

Query Match      50.0%; Score 10; DB 15; Length 363;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GFTKVCGAPP 10
      |||||
```


Db 176 GFTKVCGAPP 185

RESULT 15

US-10-187-257-4
; Sequence 4, Application US/10187257
; Publication No. US20030138458A1
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: O'HAGAN, Derek
; TITLE OF INVENTION: HCV E1E2 VACCINE COMPOSITIONS
; FILE REFERENCE: 2302-17206
; CURRENT APPLICATION NUMBER: US/10/187,257
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
US-10-187-257-4

Query Match 50.0%; Score 10; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 0.032;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPP 10

Db 387 GFTKVCGAPP 396

Search completed: November 21, 2003, 22:19:39
Job time : 19.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 163.1 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-84

Perfect score: 20

Sequence: 1 GFTKVCAPPVIGAGNNT 20

Scoring table:

Oligo Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA_Main:

- 1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/1/paa/US097A_COMB.pep.*
- 22: /cgn2_6/ptodata/1/paa/US097B_COMB.pep.*
- 23: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
- 24: /cgn2_6/ptodata/1/paa/US099A_COMB.pep.*
- 25: /cgn2_6/ptodata/1/paa/US099B_COMB.pep.*
- 26: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
- 27: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
- 28: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
- 29: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
- 30: /cgn2_6/ptodata/1/paa/US104_COMB.pep.*
- 31: /cgn2_6/ptodata/1/paa/US106_COMB.pep.*
- 32: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	13	US-08-974-690C-40
					Sequence 40, Appl

2	20	106.0	20	23	US-09-899-303-84	Sequence 84, Appl
3	20	100.0	20	23	US-09-899-303A-84	Sequence 84, Appl
4	20	100.0	20	25	US-09-973-025-84	Sequence 84, Appl
5	20	100.0	20	25	US-09-995-791-84	Sequence 84, Appl
6	20	100.0	20	25	US-09-995-808-84	Sequence 84, Appl
7	20	100.0	20	25	US-09-995-860-84	Sequence 84, Appl
8	20	100.0	20	26	US-10-020-510-84	Sequence 84, Appl
9	20	100.0	20	29	US-10-321-798-84	Sequence 84, Appl
10	10	50.0	20	13	US-08-974-685-40	Sequence 40, Appl
11	10	50.0	20	13	US-08-974-690-40	Sequence 40, Appl
12	10	50.0	20	13	US-08-974-690A-40	Sequence 40, Appl
13	10	50.0	20	13	US-08-974-690B-40	Sequence 40, Appl
14	10	50.0	26	3	US-07-667-275A-2	Sequence 2, Appl
15	10	50.0	26	8	US-08-475-482-2	Sequence 2, Appl
16	10	50.0	26	8	US-08-477-072-2	Sequence 2, Appl
17	10	50.0	26	8	US-08-477-582-2	Sequence 2, Appl
18	10	50.0	26	8	US-08-480-253-2	Sequence 2, Appl
19	10	50.0	34	8	US-08-475-482-41	Sequence 41, Appl
20	10	50.0	34	8	US-08-477-072-41	Sequence 41, Appl
21	10	50.0	34	8	US-08-477-582-41	Sequence 41, Appl
22	10	50.0	34	8	US-08-480-253-41	Sequence 41, Appl
23	10	50.0	42	8	US-08-475-482-42	Sequence 42, Appl
24	10	50.0	42	8	US-08-477-072-42	Sequence 42, Appl
25	10	50.0	42	8	US-08-477-582-42	Sequence 42, Appl
26	10	50.0	42	8	US-08-480-253-42	Sequence 42, Appl
27	10	50.0	50	8	US-08-475-482-43	Sequence 43, Appl
28	10	50.0	50	8	US-08-477-072-43	Sequence 43, Appl
29	10	50.0	50	8	US-08-477-582-43	Sequence 43, Appl
30	10	50.0	50	8	US-08-480-253-43	Sequence 43, Appl
31	10	50.0	88	8	US-08-471-498-27	Sequence 27, Appl
32	10	50.0	96	24	US-09-921-397-82	Sequence 82, Appl
33	10	50.0	112	8	US-08-436-966-18	Sequence 18, Appl
34	10	50.0	112	8	US-08-436-966-19	Sequence 19, Appl
35	10	50.0	179	24	US-09-921-397-81	Sequence 81, Appl
36	10	50.0	179	8	US-08-403-590B-77	Sequence 77, Appl
37	10	50.0	179	8	US-08-444-112-77	Sequence 77, Appl
38	10	50.0	250	1	PCT-US01-28767-8	Sequence 8, Appl
39	10	50.0	250	25	US-09-952-572-8	Sequence 8, Appl
40	10	50.0	254	1	PCT-US00-26395-3	Sequence 3, Appl
41	10	50.0	254	18	US-09-407-430-3	Sequence 3, Appl
42	10	50.0	278	8	US-08-471-498-15	Sequence 15, Appl
43	10	50.0	333	8	US-08-417-478-12	Sequence 12, Appl
44	10	50.0	333	8	US-08-453-613-12	Sequence 12, Appl
45	10	50.0	333	30	US-10-445-724-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-974-690C-40
; Sequence 40, Application US/08974690C
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-974-690C-40

```
Query Match      100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPPVCIGGAGNNT 20
Db 1 GFTKVCGAPPVCIGGAGNNT 20

RESULT 2
US-09-899-303-84
; Sequence 84, Application US/09899303
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSSE, MARIE-ANGE
;
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
;
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-899-303-84

Query Match      100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPPVCIGGAGNNT 20
Db 1 GFTKVCGAPPVCIGGAGNNT 20

RESULT 3
US-09-899-303A-84
; Sequence 84, Application US/09899303A
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSSE, MARIE-ANGE
;
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
;
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-899-303-84

Query Match      100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPPVCIGGAGNNT 20
Db 1 GFTKVCGAPPVCIGGAGNNT 20

RESULT 4
US-09-973-025-84
; Sequence 84, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSSE, MARIE-ANGE
;
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
;
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-899-303A-84

Query Match      100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPPVCIGGAGNNT 20
Db 1 GFTKVCGAPPVCIGGAGNNT 20

RESULT 5
US-09-973-025-84
; Sequence 84, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSSE, MARIE-ANGE
;
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
;
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-899-303A-84
```

```
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303A
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
;
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-899-303A-84

Query Match      100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPPVCIGGAGNNT 20
Db 1 GFTKVCGAPPVCIGGAGNNT 20

RESULT 6
US-09-973-025-84
; Sequence 84, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSSE, MARIE-ANGE
;
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
;
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-899-303A-84
```

```
Query Match      100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPPVCIGGAGNNT 20
Db 1 GFTKVCGAPPVCIGGAGNNT 20

RESULT 7
US-09-973-025-84
; Sequence 84, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSSE, MARIE-ANGE
;
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
;
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
;
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-899-303A-84
```

; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIORITY INFORMATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-973-025-84

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPPVCIGGAGNNT 20
| | | | | | | | | | | | | | | | | |
Db 1 GFTKVCGAPPVCIGGAGNNT 20

RESULT 5
US-09-995-791-84
; Sequence 84, Application US/09995791
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 84
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-84

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPPVCIGGAGNNT 20
| | | | | | | | | | | | | | | | | |
Db 1 GFTKVCGAPPVCIGGAGNNT 20

RESULT 6
US-09-995-808-84
; Sequence 84, Application US/09995808
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 84
; LENGTH: 20
; TYPE: PRT

; ORGANISM: Hepatitis C virus
US-09-995-808-84

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPPVCIGGAGNNT 20
| | | | | | | | | | | | | | | | | |
Db 1 GFTKVCGAPPVCIGGAGNNT 20

RESULT 7
US-09-995-860-84
; Sequence 84, Application US/09995860
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 84
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-84

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPPVCIGGAGNNT 20
| | | | | | | | | | | | | | | | | |
Db 1 GFTKVCGAPPVCIGGAGNNT 20

RESULT 8
US-10-020-510-84
; Sequence 84, Application US/10020510
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-72
; CURRENT APPLICATION NUMBER: US/10/020,510
; CURRENT FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 84
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-020-510-84

Query Match 100.0%; Score 20; DB 26; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPPVCIGGAGNNT 20
| | | | | | | | | | | | | | | | | |
Db 1 GFTKVCGAPPVCIGGAGNNT 20

RESULT 9
US-10-321-798-84
; Sequence 84, Application US/10321798
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.

FILE REFERENCE: 2551-93
CURRENT APPLICATION NUMBER: US/10/321,798
CURRENT FILING DATE: 2002-12-18
PRIOR APPLICATION NUMBER: 60/418,358
PRIOR FILING DATE: 2002-10-16
PRIOR APPLICATION NUMBER: 10/020,510
PRIOR FILING DATE: 2001-12-18
NUMBER OF SEQ ID NOS: 123
SOFTWARE: PatentIn 3.1
SEQ ID NO 84
LENGTH: 20
TYPE: PRT
ORGANISM: Hepatitis C virus
US-10-321-798-84

Query Match 100.0%; Score 20; DB 29; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFTKVCAPPVCIGAGNNT 20
|||||
Db 1 GFTKVCAPPVCIGAGNNT 20

RESULT 10

US-08-974-685-40

Sequence 40, Application US/08974685
GENERAL INFORMATION:
APPLICANT: LEROUIX-ROELS, GEERT
DELEYS, ROBERT
MAERTENS, GEERT

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF

HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 181

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHVE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,685

FILING DATE: 19-Nov-1997

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-11

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 40:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 40:

US-08-974-685-40

Query Match 50.0%; Score 10; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFTKVCAPP 10

Db 1 GFTKVCAPP 10
|||||

RESULT 11

US-08-974-690-40

Sequence 40, Application US/08974690

GENERAL INFORMATION:

APPLICANT: LEROUIX-ROELS, GEERT

DELEYS, ROBERT

MAERTENS, GEERT

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF

HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 166

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHVE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,690

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/635,886

FILING DATE: 25-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-11

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 40:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-974-690-40

Query Match 50.0%; Score 10; DB 13; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.018;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFTKVCAPP 10

|||||

Db 1 GFTKVCAPP 10

RESULT 12

US-08-974-690A-40

Sequence 40, Application US/08974690A

GENERAL INFORMATION:

APPLICANT: LEROUIX-ROELS, GEERT

DELEYS, ROBERT

MAERTENS, GEERT

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF

HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 177

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHVE P.C.

STREET: 1100 NORTH GLEBE ROAD

; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,690A
; FILING DATE: 19-Nov-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B. J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 2752-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 40:
; Best Local Similarity 50.0%; Score 10; DB 13; Length 20;
; Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-08-974-690A-40

Query Match 50.0%; Score 10; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GFTKVCGAPP 10
Db 1 GFTKVCGAPP 10
|||||
RESULT 13
US-08-974-690B-40
; Sequence 40, Application US/08974690B
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; DELEYS, ROBERT
; MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
; THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,690B
; FILING DATE: 19-Nov-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B. J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 2752-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100

; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-08-974-690B-40

Query Match 50.0%; Score 10; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GFTKVCGAPP 10
Db 1 GFTKVCGAPP 10
|||||
RESULT 14 0
US-07-667-275A-2
; Sequence 2, Application US/07667275A
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; APPLICANT: Hosein, Barbara M.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR THE
; DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV INFECTION
; AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/07/667,275A
; APPLICATION NUMBER: 514
; FILING DATE: 19910311
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/651,735
; FILING DATE: 07-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/558,799
; FILING DATE: 26-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/510,153
; FILING DATE: 16-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/481/348
; FILING DATE: 16-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Lin, Maria C.H.
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 11514043B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: 212-751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

; ORIGINAL SOURCE:
; ORGANISM: Hepatitis C Virus
US-07-667-275A-2

Query Match 50.0%; Score 10; DB 3; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFTKVCGAPP 10
Db 16 GFTKVCGAPP 25

RESULT 15

US-08-475-482-2
; Sequence 2, Application US/08475482
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,482
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 17-June-1994
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US10
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-475-482-2

Query Match 50.0%; Score 10; DB 8; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFTKVCGAPP 10
Db 16 GFTKVCGAPP 25

Db 16 GFTKVCGAPP 25

Search completed: November 21, 2003, 22:09:52
Job time : 163.1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.55 Seconds
(without alignments)
117.536 Million cell updates/sec

Title: US-09-973-025-84

Perfect score: 20
Sequence: 1 GFTKVCAPPVCIGGAGNNT 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents AA, New.*
1: /cgn2_6/ptodata/1/paa/pct_NEW_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	6	US-10-651-165-40
2	10	50.0	30	6	US-10-296-734-480
3	10	50.0	30	6	US-10-296-734-482
4	10	50.0	333	1	PCT-US03-19834-3
5	10	50.0	333	6	US-10-655-562-4
6	10	50.0	637	1	PCT-US03-33610-4
7	10	50.0	2010	6	US-10-296-734-814
8	10	50.0	2011	6	US-10-296-734-812
9	10	50.0	3011	1	PCT-US03-19834-2
10	10	50.0	3011	6	US-10-296-734-406
11	10	50.0	5985	6	US-10-296-734-810
12	8	40.0	10	6	US-10-440-390-35
13	8	40.0	20	6	US-10-651-165-39
14	8	40.0	20	6	US-10-651-165-41
15	8	40.0	30	6	US-10-296-734-484
16	8	40.0	30	6	US-10-685-435-26
17	8	40.0	68	6	US-10-651-165-250
18	8	40.0	68	6	US-10-651-165-260
19	8	40.0	347	6	US-10-664-391-9
20	8	40.0	539	6	US-10-664-391-11
21	6	30.0	64	5	US-09-882-583A-14
22	6	30.0	357	6	US-10-679-063-1523
23	6	30.0	515	6	US-10-679-063-15436
24	5	25.0	9	6	US-10-428-335-55
25	5	25.0	9	6	US-10-428-335-58
26	5	25.0	9	6	US-10-428-335-71

27	5	25.0	9	6	US-10-428-335-97
28	5	25.0	24	1	PCT-US03-06553-9
29	5	25.0	24	1	PCT-US03-20684-24
30	5	25.0	24	5	US-09-862-179C-37
31	5	25.0	24	6	US-10-275-427A-42
32	5	25.0	64	5	US-09-882-583A-9
33	5	25.0	64	5	US-09-882-583A-10
34	5	25.0	64	5	US-09-882-583A-13
35	5	25.0	72	5	US-09-897-516A-5225
36	5	25.0	84	6	US-10-473-305-1293
37	5	25.0	85	5	US-09-897-516A-5821
38	5	25.0	97	7	US-60-500-315-1475
39	5	25.0	100	6	US-10-425-114A-53005
40	5	25.0	109	6	US-10-659-199-16
41	5	25.0	133	6	US-10-425-114A-48117
42	5	25.0	133	6	US-10-425-114A-59096
43	5	25.0	144	6	US-10-425-114A-56848
44	5	25.0	150	7	US-60-505-948-26
45	5	25.0	164	7	US-60-495-114-1808

ALIGNMENTS

RESULT 1
US-10-651-165-40
; Sequence 40, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-40

Query Match 100.0%; Score 20; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPPVCIGGAGNNT 20
Db 1 GFTKVCAPPVCIGGAGNNT 20
|||||
|||||

RESULT 2
US-10-296-734-480
; Sequence 480, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 480
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 37

US-10-296-734-480

Query Match 50.0%; Score 10; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.0033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPP 10
| | | | | | | | | |
Db 21 GFTKVCGAPP 30

RESULT 3

US-10-296-734-482
; Sequence 482, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 482
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 38

US-10-296-734-482

Query Match 50.0%; Score 10; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.0033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPP 10
| | | | | | | | | |
Db 6 GFTKVCGAPP 15

RESULT 4

PCT-US03-19834-3
; Sequence 3, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; FILE REFERENCE: IOWA:045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Hepatitis C virus

PCT-US03-19834-3

Query Match 50.0%; Score 10; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 0.027;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GFTKVCGAPP 10
| | | | | | | | | |
Db 177 GFTKVCGAPP 186.

RESULT 5

US-10-655-562-4
; Sequence 4, Application US/10655562
; GENERAL INFORMATION:
; APPLICANT: BISHOP-HURLEY, SHARON L.
; APPLICANT: SCHMIDT, FRANCIS J.
; APPLICANT: SMITH, ARNOLD L.
; TITLE OF INVENTION: PHAGE-DISPLAY PEPTIDES AS NOVEL ANTIMICROBIAL AGENTS
; FILE REFERENCE: UVMO:022US
; CURRENT APPLICATION NUMBER: US/10/655,562
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: 60,409,909
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-655-562-4

Query Match 50.0%; Score 10; DB 6; Length 333;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPP 10
| | | | | | | | | |
Db 177 GFTKVCGAPP 186

RESULT 6

PCT-US03-33610-4
; Sequence 4, Application PC/TUS0333610
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: SELBY, Mark
; APPLICANT: PALIARD, Xavier
; TITLE OF INVENTION: ACTIVATION OF HCV-SPECIFIC T-CELLS
; FILE REFERENCE: #300-1612.60
; CURRENT APPLICATION NUMBER: PCT/US03/33610
; CURRENT FILING DATE: 2003-10-24
; PRIOR APPLICATION NUMBER: 10/281,341
; PRIOR FILING DATE: 2002-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
PCT-US03-33610-4

Query Match 50.0%; Score 10; DB 1; Length 637;
Best Local Similarity 100.0%; Pred. No. 0.048;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPP 10
| | | | | | | | | |
Db 387 GFTKVCGAPP 396

```
RESULT 7
US-10-296-734-814
; Sequence 814, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 814
; LENGTH: 2010
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette B
US-10-296-734-814

Query Match          50.0%; Score 10; DB 6; Length 2010;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GFTKVCGAPP 10
Db      396 GFTKVCGAPP 405

RESULT 8
US-10-296-734-812
; Sequence 812, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 812
; LENGTH: 2011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette A
US-10-296-734-812

Query Match          50.0%; Score 10; DB 6; Length 2011;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GFTKVCGAPP 10
Db      412 GFTKVCGAPP 421

RESULT 9
PCT-US03-19834-2
; Sequence 2, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; FILE REFERENCE: IOWA:045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834

Query Match          50.0%; Score 10; DB 6; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GFTKVCGAPP 10
Db      559 GFTKVCGAPP 568

RESULT 10
US-10-296-734-406
; Sequence 406, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la consensus polyprotein
US-10-296-734-406

Query Match          50.0%; Score 10; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GFTKVCGAPP 10
Db      559 GFTKVCGAPP 568

RESULT 11
US-10-296-734-810
; Sequence 810, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 810
; LENGTH: 5985
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la savine
```

US-10-296-734-810

Query Match 50.0%; Score 10; DB 6; Length 5985;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFTKVCGAPP 10
|||||

DB 411 GFTKVCGAPP 420

RESULT 12

US-10-440-390-35
; Sequence 35, Application US/10440390
; GENERAL INFORMATION:
; APPLICANT: Laufer et al.
; TITLE OF INVENTION: Epitopes of Hepatitis C Virus
; FILE REFERENCE: 24028-011
; CURRENT APPLICATION NUMBER: US/10/440,390
; PRIOR FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: 60/381,273
; PRIOR FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-440-390-35

Query Match 40.0%; Score 8; DB 6; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 IGGAGNNT 20
|||||

DB 3 IGGAGNNT 10

RESULT 13

US-10-651-165-39
; Sequence 39, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-39

Query Match 40.0%; Score 8; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFTKVCGA 8
|||||

DB 13 GFTKVCGA 20

RESULT 14

US-10-651-165-41
; Sequence 41, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-41

Query Match 40.0%; Score 8; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 IGGAGNNT 20
|||||

DB 1 IGGAGNNT 8

RESULT 15

US-10-296-734-484
; Sequence 484, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 484
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 39
US-10-296-734-484

Query Match 40.0%; Score 8; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 IGGAGNNT 20
|||||

DB 3 IGGAGNNT 10

Search completed: November 21, 2003, 22:12:57
Job time : 8.55 secs

Wed Nov 26 17:25:40 2003

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:46:27 ; Search time 9.25 Seconds
(without alignments)
207.932 Million cell updates/sec

Title: US-09-973-025-84

Perfect score: 20

Sequence: 1 GFTKVCGAPPVCIGAGNNT 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR 76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	50.0	640	JQ1584	genome polyprotein
2	10	50.0	716	JQ1366	polyprotein - hepa
3	10	50.0	3011	GNWVC3	genome polyprotein
4	10	50.0	3011	GNWVCH	genome polyprotein
5	10	50.0	3011	S40770	genome polyprotein
6	8	40.0	782	S19876	genome polyprotein
7	8	40.0	782	S19875	genome polyprotein
8	8	40.0	3010	S18030	genome polyprotein
9	7	35.0	364	B64346	cell division prot
10	7	35.0	369	G64224	cell division prot
11	7	35.0	375	T44848	cell division prot
12	7	35.0	375	H84179	cell division prot
13	7	35.0	380	S73845	cell division prot
14	7	35.0	381	A63091	cell division prot
15	7	35.0	398	H70393	hemolysin - Aquife
16	7	35.0	403	F64377	cell division prot
17	7	35.0	662	S55274	alpha-N-arabinofur
18	7	35.0	1023	LESCA	hemolysin A - Esch
19	7	35.0	3262	AH2137	hypothetical prote
20	6	30.0	174	D70979	probable purE prot
21	6	30.0	257	S74478	hypothetical prote
22	6	30.0	313	H96037	alpha-galactoside
23	6	30.0	351	H72328	cell division prot
24	6	30.0	371	E75494	cell division prot
25	6	30.0	372	AE1367	carbohydrate kinas
26	6	30.0	372	AD1736	carbohydrate kinas
27	6	30.0	377	AD1937	permease protein o
28	6	30.0	386	A96532	hypothetical prote
29	6	30.0	427	F64064	to1B protein - Hae

30 6 30.0 444 2 S40980
31 6 30.0 446 1 UBZPG
32 6 30.0 449 2 H69423
33 6 30.0 451 1 UBHUG
34 6 30.0 451 1 UBXLG
35 6 30.0 455 2 G71860
36 6 30.0 457 2 T08419
37 6 30.0 458 2 B64566
38 6 30.0 462 2 S53084
39 6 30.0 468 2 T08057
40 6 30.0 468 2 T07904
41 6 30.0 469 2 S31727
42 6 30.0 469 2 S44193
43 6 30.0 472 2 S39553
44 6 30.0 474 2 T50558
45 6 30.0 474 2 T47957

ALIGNMENTS

RESULT 1

JQ1584

genome polyprotein - hepatitis C virus (strain U.K.) (fragment)

N:Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural prot
C:Species: hepatitis C virus

C>Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000

C:Accession: JQ1584

R:Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.

J. Gen. Virol. 73, 1521-1525, 1992

A:Title: Cloning and sequencing of the structural region and expression of putative core

A:Reference number: JQ1584; MUID:92300349; PMID:1318944

A:Accession: JQ1584

A:Molecule type: genomic RNA

A:Residues: 1-840 <KUM>

A:Cross-references: GB:X84079; NID:G643119; PIDN:CAA58888.1; PID:G643120

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; polypro

F:1-191/Product: core protein C #status predicted <CPC>

F:192-389/Product: envelope protein E1 #status predicted <EEL>

F:390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predicted <

F:196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn) (cov

Query Match 50.0%; Score 10; DB 2; Length 640;

Best Local Similarity 100.0%; Pred. No. 0.0057;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCGAPP 10

Db 559 GFTKVCGAPP 568

RESULT 2

JQ1366

polyprotein^o- hepatitis C virus (French isolate) (fragments)

C:Species: hepatitis C virus

C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000

C:Accession: JQ1366

R:Kremsdorff, D.; Forchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.

J. Gen. Virol. 72, 2557-2561, 1991

A:Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication

A:Reference number: JQ1366; MUID:92013977; PMID:1655961

A:Accession: JQ1366

A:Molecule type: genomic RNA

A:Residues: 1-716 <KRE>

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: glycoprotein; polyprotein

F:84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #stat

Query Match

Best Local Similarity 100.0%; Score 10; DB 2; Length 716;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

A:Accession: A36914
A:Molecule type: Genomic RNA
A:Residues: 1-3011 <INC>
A:Cross-references: GB:M67463; NID:G329737; PIDN:AAA45534.1; PID:G329738
R:Inchauspe, G.; Zebede, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: comparison with the Japanese isolate H-77C.
A:Reference number: A41546; MUID:92052256; PMID:1659800
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis virus #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,2303,2304,2305,2306,2307,2308,2309,2310,2311,2312,2313,2314,2315,2316,2317,2318,2319,2320,2321,2322,2323,2324,2325,2326,2327,2328,2329,2330,2331,2332,2333,2334,2335,2336,2337,2338,2339,2340,2341,2342,2343,2344,2345,2346,2347,2348,2349,2350,2351,2352,2353,2354,2355,2356,2357,2358,2359,2360,2361,2362,2363,2364,2365,2366,2367,2368,2369,2370,2371,2372,2373,2374,2375,2376,2377,2378,2379,2380,2381,2382,2383,2384,2385,2386,2387,2388,2389,2390,2391,2392,2393,2394,2395,2396,2397,2398,2399,2400,2401,2402,2403,2404,2405,2406,2407,2408,2409,2410,2411,2412,2413,2414,2415,2416,2417,2418,2419,2420,2421,2422,2423,2424,2425,2426,2427,2428,2429,2430,2431,2432,2433,2434,2435,2436,2437,2438,2439,2440,2441,2442,2443,2444,2445,2446,2447,2448,2449,2450,2451,2452,2453,2454,2455,2456,2457,2458,2459,2460,2461,2462,2463,2464,2465,2466,2467,2468,2469,2470,2471,2472,2473,2474,2475,2476,2477,2478,2479,2480,2481,2482,2483,2484,2485,2486,2487,2488,2489,2490,2491,2492,2493,2494,2495,2496,2497,2498,2499,2500,2501,2502,2503,2504,2505,2506,2507,2508,2509,2510,2511,2512,2513,2514,2515,2516,2517,2518,2519,2520,2521,2522,2523,2524,2525,2526,2527,2528,2529,2530,2531,2532,2533,2534,2535,2536,2537,2538,2539,2540,2541,2542,2543,2544,2545,2546,2547,2548,2549,2550,2551,2552,2553,2554,2555,2556,2557,2558,2559,2560,2561,2562,2563,2564,2565,2566,2567,2568,2569,2570,2571,2572,2573,2574,2575,2576,2577,2578,2579,2580,2581,2582,2583,2584,2585,2586,2587,2588,2589,2590,2591,2592,2593,2594,2595,2596,2597,2598,2599,2600,2601,2602,2603,2604,2605,2606,2607,2608,2609,2610,2611,2612,2613,2614,2615,2616,2617,2618,2619,2620,2621,2622,2623,2624,2625,2626,2627,2628,2629,2630,2631,2632,2633,2634,2635,2636,2637,2638,2639,2640,2641,2642,2643,2644,2645,2646,2647,2648,2649,2650,2651,2652,2653,2654,2655,2656,2657,2658,2659,2660,2661,2662,2663,2664,2665,2666,2667,2668,2669,2670,2671,2672,2673,2674,2675,2676,2677,2678,2679,2680,2681,2682,2683,2684,2685,2686,2687,2688,2689,2690,2691,2692,2693,2694,2695,2696,2697,2698,2699,2700,2701,2702,2703,2704,2705,2706,2707,2708,2709,2710,2711,2712,2713,2714,2715,2716,2717,2718,2719,2720,2721,2722,2723,2724,2725,2726,2727,2728,2729,2730,2731,2732,2733,2734,2735,2736,2737,2738,2739,2740,2741,2742,2743,2744,2745,2746,2747,2748,2749,2750,2751,2752,2753,2754,2755,2756,2757,2758,2759,2760,2761,2762,2763,2764,2765,2766,2767,2768,2769,2770,2771,2772,2773,2774,2775,2776,2777,2778,2779,2780,2781,2782,2783,2784,2785,2786,2787,2788,2789,2790,2791,2792,2793,2794,2795,2796,2797,2798,2799,2800,2801,2802,2803,2804,2805,2806,2807,2808,2809,2810,2811,2812,2813,2814,2815,2816,2817,2818,2819,2820,2821,2822,2823,2824,2825,2826,2827,2828,2829,2830,2831,2832,2833,2834,2835,2836,2837,2838,2839,2840,2841,2842,2843,2844,2845,2846,2847,2848,2849,2850,2851,2852,2853,2854,2855,2856,2857,2858,2859,2860,2861,2862,2863,2864,2865,2866,2867,2868,2869,2870,2871,2872,2873,2874,2875,2876,2877,2878,2879,2880,2881,2882,2883,2884,2885,2886,2887,2888,2889,2890,2891,2892,2893,2894,2895,2896,2897,2898,2899,2900,2901,2902,2903,2904,2905,2906,2907,2908,2909,2910,2911,2912,2913,2914,2915,2916,2917,2918,2919,2920,2921,2922,2923,2924,2925,2926,2927,2928,2929,2930,2931,2932,2933,2934,2935,2936,2937,2938,2939,2940,2941,2942,2943,2944,2945,2946,2947,2948,2949,2950,2951,2952,2953,2954,2955,2956,2957,2958,2959,2960,2961,2962,2963,2964,2965,2966,2967,2968,2969,2970,2971,2972,2973,2974,2975,2976,2977,29

Query Match 50.0%; Score 10; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 0.02;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFTKVCAPP 10
 |||||
 DB 559 GFTKVCAPP 568

RESULT 6
 S19876
 genome polypeptide - hepatitis C virus (isolate JK5) (fragment)
 N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
 C:Species: hepatitis C virus
 A:Variety: isolate JK5
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
 C:Accession: S19876
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
 A:Reference number: S18029
 A:Accession: S19876
 A:Molecule type: genomic RNA
 A:Residues: 1-782 <HON>
 A:Cross-references: EMBL:X61595; NID:g59486; PIDN:CAA43792.1; PID:g59487
 A:Experimental source: isolate JK5
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
 F1-191/Product: core protein #status predicted <MAT1>
 F192-383/Product: envelope protein 1 #status predicted <MAT2>
 F384-733/Product: NS1/E2 protein #status predicted <MAT3>
 F734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 40.0%; Score 8; DB 2; Length 782;
 Best Local Similarity 100.0%; Pred. No. 0.83;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 IGGAGNNT 20
 |||||
 DB 571 IGGAGNNT 578

RESULT 7
 S19875
 genome polypeptide - hepatitis C virus (isolate JK3) (fragment)
 N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
 C:Species: hepatitis C virus
 A:Variety: isolate JK3
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
 C:Accession: S19875
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
 A:Reference number: S18029
 A:Accession: S19875
 A:Molecule type: genomic RNA
 A:Residues: 1-782 <HON>
 A:Cross-references: EMBL:X61592; NID:g59482; PIDN:CAA43789.1; PID:g59483
 A:Experimental source: isolate JK3
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
 F1-191/Product: core protein #status predicted <MAT1>
 F192-383/Product: envelope protein 1 #status predicted <MAT2>
 F384-733/Product: NS1/E2 protein #status predicted <MAT3>
 F734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 40.0%; Score 8; DB 2; Length 782;
 Best Local Similarity 100.0%; Pred. No. 0.83;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 IGGAGNNT 20
 |||||

Db 571 IGGAGNNT 578

RESULT 8
 S18030
 genome polypeptide - hepatitis C virus (isolate JK1)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 A:Variety: isolate JK1
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
 C:Accession: S18030; S33570; A48332; S18029
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single patie
 A:Reference number: S18028
 A:Accession: S18030
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <HON>
 A:Cross-references: EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59479
 A:Experimental source: isolate JK1 from an individual
 R:Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.
 Arch. Virol. 128, 163-169, 1993
 A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolated
 A:Reference number: A48332; MUID:93119270; PMID:8380322
 A:Accession: S33570
 A:Molecule type: genomic RNA
 A:Residues: 1-547, 'T', 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HOW>
 A:Cross-references: EMBL:X61591
 A:Note: this sequence is inconsistent with the nucleotide translation
 A:Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320
 as Trp, and TTC for residue 771 as Ser
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin
 F12-115/Product: capsid protein C #status predicted <CPC>
 F116-191/Product: envelope protein M #status predicted <EPM>
 F192-389/Product: major envelope protein E #status predicted <MEE>
 F390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F1007-1615/Product: hepatitis C virus genome polypeptide
 F1230-1237/Region: nucleotide-binding motif A (P-loop)
 F1312-1317/Region: nucleotide-binding motif B
 F1316-1319/Region: DEXH motif
 F1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F2034-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F196, 209, 234, 250, 305, 417, 423, 448, 532, 540, 556, 576, 623, 645/Binding site: carbohydrate (AS

Query Match 40.0%; Score 8; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 2.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 IGGAGNNT 20
 |||||
 Db 571 IGGAGNNT 578

RESULT 9
 B64346
 cell division protein FtsZ homolog MJ0370 - Methanococcus jannaschii
 C:Species: Methanococcus jannaschii
 C:Date: 13-Sep-1996 #sequence_revision 04-Oct-1996 #text_change 21-Jul-2000
 C:Accession: B64346
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
 Reisch, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
 reon, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273, 1058-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
 A:Reference number: A64300; MUID:96337999; PMID:8688087
 A:Accession: B64346
 A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-364 <BUL>
A:CROSS-references: GB:U67490; GB:L77117; NID:g2826276; PIDN:AAB98359.1; PID:g1591077; T
C:Comment: This protein is the nearer of two homologs in the archaeon Methanococcus jann
f self-assembly and required for initiating the partition between compartments during ce
C:Genetics:
A:Map position: FOR336314-337408
A:Start codon: GTG
C:Superfamily: cell division protein ftsZ
C:Keywords: cell division; GTP binding
F:132-138/Region: tubulin/FtsZ GTP/GDP-binding (G-G-G-T-G-[ST]-G) motif

Query Match 35.0%; Score 7; DB 1; Length 364;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 GGAGNNT 20
Db 46 GGAGNNT 52

RESULT 10
G64224
cell division protein FtsZ - Mycoplasma genitalium
C:Species: Mycoplasma genitalium
C:Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 07-Dec-1999
R:Praser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.
C.A.; Venter, J.C.
Science 270, 397-403, 1995
A:Title: The minimal gene complement of Mycoplasma genitalium.
A:Reference number: A64200; MUID:96026346; PMID:7569993
A:Accession: G64224
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-369 <TIGR>
A:CROSS-references: GB:U93700; GB:L43967; NID:g1045906; PID:g1045912; TIGR:MG224
A:Experimental source: strain G-37
C:Genetics:
A:Genetic code: SGC3
C:Superfamily: cell division protein ftsZ
C:Keywords: cell division; GTP binding
F:117-123/Region: tubulin/FtsZ GTP/GDP-binding (G-G-G-T-G-[ST]-G) motif #status atypical

Query Match 35.0%; Score 7; DB 2; Length 369;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 IGGAGNN 19
Db 25 IGGAGNN 31

RESULT 11
T44848
cell division protein ftsZ [validated] - Halobacterium salinarum
N:Alternate names: GTP-binding protein ftsZ
C:Species: Halobacterium salinarum
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T44848
R:Margolin, W.; Wang, R.; Kumar, M.
J. Bacteriol. 178, 1320-1327, 1996
A:Title: Isolation of an ftsZ homolog from the archaeobacterium Halobacterium salinarum:
A:Reference number: Z22858; MUID:96200101; PMID:8631708
A:Accession: T44848
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-375 <MAR>
A:CROSS-references: EMBL:U32860; PIDN:AAB06191.1
C:Genetics:
A:Gene: ftsZ
C:Function:

A:Description: involved in cell division; overexpression of ftsZ induces significant mor
C:Superfamily: cell division protein ftsZ
C:Keywords: cell division; GTP binding

Query Match 35.0%; Score 7; DB 1; Length 375;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 GGAGNNT 20
Db 21 GGAGNNT 27

RESULT 12
H84179
cell division protein [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: H84179
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabl
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: H84179
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-375 <STO>
A:CROSS-references: GB:AE004437; NID:g10579938; PIDN:AAG18804.1; GSPDB:GN00138
C:Genetics:
A:Gene: ftsZ
C:Superfamily: cell division protein ftsZ

Query Match 35.0%; Score 7; DB 2; Length 375;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 GGAGNNT 20
Db 21 GGAGNNT 27

RESULT 13
S73845
cell division protein ftsZ - Mycoplasma pneumoniae (strain ATCC 29342)
N:Alternate names: hypothetical protein F10_orf380
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
C:Accession: S73845
R:Himmelfreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73327; MUID:97105885; PMID:8948633
A:Accession: S73845
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-380 <HIM>
A:CROSS-references: EMBL:AE000051; GB:U00089; NID:g1674211; PIDN:AAB96167.1; PID:g167421
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Gene: ftsZ
A:Genetic code: SGC3
C:Superfamily: cell division protein ftsZ
F:117-123/Region: tubulin/FtsZ GTP/GDP-binding (G-G-G-T-G-[ST]-G) motif #status atypical

Query Match 35.0%; Score 7; DB 2; Length 380;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 IGGAGNN 19


```
Db          25 IGGAGNN 31
|||||
RESULT 14
A69091
cell division protein FtsZ - Methanobacterium thermoautotrophicum (strain Delta H)
C:Species: Methanobacterium thermoautotrophicum
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-Aug-1999
C:Accession: A69091
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: func
A:Reference number: A69000; MUID:98037514; PMID:9371463
A:Accession: A69091
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-381 <MTH>
A:Cross-references: GB:AE000925; GB:AE000666; NID:g2622791; PIDN:AAB86148.1; PID:g262280
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1676
A:Start codon: TTG
C:Superfamily: cell division protein ftsZ
F:133-139/Region: tubulin/FtsZ GTP/GDP-binding (G-G-G-T-G-[ST]-G) motif
Query Match          35.0%; Score 7; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 14 GGAGNNT 20
|||||
Db 47 GGAGNNT 53
|||||
RESULT 15
H70393
hemolysin - Aquifex aeolicus
C:Species: Aquifex aeolicus
C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C:Accession: H70393
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: H70393
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-398 <AQF>
A:Cross-references: GB:AE000722; NID:g2983559; PIDN:AAC07135.1; PID:g2983562; GB:AE00065
A:Experimental source: strain VF5
C:Genetics:
A:Gene: hylA
Query Match          35.0%; Score 7; DB 2; Length 398;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 14 GGAGNNT 20
|||||
Db 159 GGAGNNT 165
|||||
Search completed: November 21, 2003, 21:11:35
Job time : 10.25 secs
```

THIS PAGE BLANK (U)SPTO

DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRp.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR007095; RNA_pol_DS_Ps.
 DR InterPro; IPR007094; RNA_pol_PsVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PDI86062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).
 FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 3012 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1093 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 50.0%; Score 10; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 0.0078;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFTKVCAPP 10

Db 559 GFTKVCAPP 568
 RESULT 2
 POLG_HCVH STANDARD; PRT; 3011 AA.
 ID POLG_HCVH
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 (GP68) (GP70); Protein P7; Nonstructural protein NS2 (P21)
 (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=1658800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 RL Structure 6:89-100(1998).
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: Nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC EMBL; M67463; AAA45534.1; -
DR PIR; A36814; GNWVCH.
DR PDB; 1HEI; 25-NOV-98.
DR PDB; 1AIV; 16-FEB-99.
DR PDB; 1A1R; 17-JUN-98.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR TRANSFAC; T04155; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA pol DS ps.
DR InterPro; IPR007094; RNA pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00489; DEXDC; 1.
KW Polypeptin; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
FT CHAIN 12 383 ENVELOPE GLYCOPROTEIN E1.
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E2.
FT CHAIN 747 809 PROTEIN P7.
FT CHAIN 810 1026 NONSTRUCTURAL PROTEIN NS2.
FT CHAIN 1027 1657 PROTEASE/HELICASE NS3.
FT CHAIN 1658 1711 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS5A.
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5B.
FT CHAIN 3011 369 POTENTIAL.
FT TRANSMEM 347 369 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1083 1093 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DESCH BOX.
FT CARBOHYD 186 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).

FT STRAND 1224 1226
FT TURN 1232 1233
FT TURN 1236 1238
FT HELIX 1239 1246
FT TURN 1247 1248
FT STRAND 1251 1255
FT HELIX 1258 1271
FT TURN 1272 1272
FT STRAND 1277 1280
FT TURN 1281 1282
FT STRAND 1283 1285
FT STRAND 1291 1295
FT HELIX 1296 1301
FT TURN 1302 1303
FT STRAND 1312 1316
FT TURN 1317 1319
FT HELIX 1323 1335
FT TURN 1336 1340
FT STRAND 1343 1347
FT TURN 1352 1353
FT STRAND 1360 1361
FT TURN 1362 1366
FT STRAND 1368 1368
FT STRAND 1373 1375
FT TURN 1376 1377
FT STRAND 1378 1380
FT HELIX 1382 1385
FT STRAND 1389 1393
FT HELIX 1397 1409
FT TURN 1410 1411
FT STRAND 1414 1417
FT TURN 1419 1420
FT STRAND 1432 1436
FT TURN 1438 1439
FT STRAND 1450 1453
FT STRAND 1456 1463
FT STRAND 1471 1478
FT STRAND 1480 1480
FT HELIX 1481 1488
FT TURN 1489 1490
FT STRAND 1497 1501
FT STRAND 1507 1507
FT STRAND 1511 1511
FT HELIX 1514 1527
FT STRAND 1532 1544
FT STRAND 1550 1550
FT HELIX 1555 1564
FT HELIX 1570 1578
FT TURN 1579 1580
FT TURN 1584 1597
FT TURN 1598 1598
FT HELIX 1606 1611
FT TURN 1614 1618
FT STRAND 1622 1623
FT STRAND 1627 1627
FT STRAND 1635 1636
FT HELIX 1640 1652
SQ SEQUENCE 3011 AA; 772CBB29CCD94753 CRC64;
Query Match 50.0%; Score 10; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.0078;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GFTKYVGAPP 10
Db 589 GFTKYVGAPP 568
RESULT 3
FTZ1 METJA STANDARD; PRT; 364 AA.
ID FTZ1_METJA
AC Q57816;
DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cell division protein ftsz homolog 1.
 GN FTSZ1 OR MJ0370.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.P., Fuhrmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts C.M., Hurd M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschii.";
 RL Science 273:1058-1073(1996).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RX MEDLINE=98089046; PubMed=9428770;
 RA Loewe J., Amos L.A.;
 RT "Crystal structure of the bacterial cell-division protein ftsz.";
 RL Nature 391:203-206(1998).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 23-356.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=99160450; PubMed=10049809;
 RA Loewe J.;
 RT "Crystal structure determination of FtsZ from Methanococcus
 RT jannaschii.";
 RL J. Struct. Biol. 124:235-243(1998).
 CC -!- FUNCTION: This protein is essential to the cell-division process.
 CC It seems to assemble into a dynamic ring on the inner surface of
 CC the cytoplasmic membrane at the place where division will occur,
 CC and the formation of the ring is the signal for septation to
 CC begin. Binds to and hydrolyzes GTP (By similarity).
 CC -!- SUBUNIT: Aggregates to form a ring-like structure (By similarity).
 CC -!- DOMAIN: CONSISTS OF TWO GLOBULAR DOMAINS WITH A CONNECTING CORE
 CC HELIX.
 CC -!- SIMILARITY: BELONGS TO THE FTSZ FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U67490; AAB98359.1; -.
 DR PIR; B64346; B64346.
 DR PDB; 1FSZ; 24-JUN-98.
 DR TIGR; MJ0370; -.
 DR InterPro; IPR000158; FtsZ.
 DR InterPro; IPR003008; Tubulin_FtsZ.
 DR Pfam; PF00091; tubulin; 1.
 DR Pfam; PF03953; tubulin C; 1.
 DR PRINTS; PR00423; CELLDIVFTSZ.
 DR TIGRFAMS; TIGR00065; ftsz; 1.
 DR PROSITE; PS01134; FTSZ_1; 1.
 DR PROSITE; PS01135; FTSZ_2; 1.
 DR Cell division; Septation; GTP-binding; Multigene family; 3D-structure;
 KW Complete proteome.
 FT NP BIND 130 138 GTP.
 FT HELIX 24 34
 FT STRAND 40 45
 FT HELIX 46 59

FT TURN 62 63
 FT STRAND 65 70
 FT STRAND 72 72
 FT HELIX 73 77
 FT TURN 78 78
 FT STRAND 83 86
 FT TURN 89 91
 FT STRAND 94 95
 FT TURN 98 99
 FT TURN 101 110
 FT TURN 111 111
 FT HELIX 112 119
 FT TURN 120 121
 FT STRAND 124 130
 FT TURN 131 132
 FT HELIX 135 149
 FT TURN 150 151
 FT STRAND 153 160
 FT HELIX 163 165
 FT HELIX 167 182
 FT TURN 183 183
 FT STRAND 186 191
 FT HELIX 192 197
 FT TURN 198 198
 FT HELIX 204 227
 FT HELIX 236 243
 FT TURN 244 245
 FT STRAND 248 256
 FT HELIX 261 271
 FT TURN 273 274
 FT TURN 279 280
 FT STRAND 285 291
 FT TURN 293 294
 FT HELIX 297 308
 FT TURN 309 310
 FT TURN 313 314
 FT STRAND 317 323
 FT TURN 328 329
 FT STRAND 331 337
 FT TURN 341 343
 FT STRAND 344 347
 FT TURN 348 349
 FT STRAND 350 353
 SQ SEQUENCE 364 AA; 38924 MW; 3BB386A5D2FCA107 CRC64;

Query Match 35.0%; Score 7; DB 1; Length 364;
 Best Local Similarity 100.0%; Pred.No. 2;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 GGAGNNT 20
 |||||
 Db 46 GGAGNNT 52

RESULT 4
 FTSZ_MYCGE STANDARD; PRT; 369 AA.
 ID FTSZ_MYCGE
 AC P47466;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cell division protein ftsz.
 GN FtsZ OR MG224.
 OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE=96026346; PubMed=7569993;
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Pritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,


```
CC CC -!- SUBUNIT: Aggregates to form a ring-like structure (By similarity).
CC CC -!- DOMAIN: CONSISTS OF TWO GLOBULAR DOMAINS WITH A CONNECTING CORE
CC CC HELIX (BY SIMILARITY).
CC CC -!- SIMILARITY: BELONGS TO THE FTSZ FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to licensese@isb-sib.ch)
CC CC -----
CC CC EMBL; U67510; AAC98617.1; -.
CC DR HSSP; Q57816; 1FSZ.
CC DR TIGR; MJ0622; -.
CC DR InterPro; IPR000158; FtsZ.
CC DR InterPro; IPR003008; Tubulin_FtsZ.
CC DR Pfam; PF00091; tubulin_1.
CC DR Pfam; PF03953; tubulin_C1.
CC DR PRINTS; PRO0423; CELLDVISFTSZ.
CC DR TIGRFAMs; TIGR00065; ftsz; 1.
CC DR PROSITE; PS01134; FTSZ_1; 1.
CC DR PROSITE; PS01135; FTSZ_2; 1.
CC DR Cell division; Septation; GTP-binding; Multigene family;
CC KW Complete proteome.
CC NP BIND 130 138 GTP (POTENTIAL).
CC FT SEQUENCE 380 AA; 40658 MW; 2A3AD01CCB54PB8C CRC64;
CC SQ -----
Query Match 35.0%; Score 7; DB 1; Length 380;
Best Localo%similarity 100.0%; Pred.No.2.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 14 GGAGNNT 20
DB 46 GGAGNNT 52
|||||
|||
RESULT 8
FTSZ METH STANDARD; PRT; 381 AA.
AC AC 02712;
DT DT 15-JUL-1998 (Rel. 36, Created)
DT DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE DE Cell division protein ftsz.
GN GN FTSZ OR MTH1676.
OS OS Methanobacterium thermoautotrophicum.
OC OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OX OX Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
[1] _TaxID=187420;
RN RP SEQUENCE FROM N.A.
RC RC STRAIN=Delta H;
RX RX MEDLINE=98037514; PubMed=9371463;
RA RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang R., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwan N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT delatH: functional analysis and comparative genomics.";
RL RL J. Bacteriol. 179:7155-7155(1997).
CC CC -!- FUNCTION: This protein is essential to the cell-division process.
CC CC It seems to assemble into a dynamic ring on the inner surface of
CC CC the cytoplasmic membrane at the place where division will occur,
CC CC and the formation of the ring is the signal for septation to
CC CC begin. Binds to and hydrolyzes GTP (By similarity)
CC CC -!- SUBUNIT: Aggregates to form a ring-like structure (By similarity).
CC CC -!- SIMILARITY: BELONGS TO THE FTSZ FAMILY.
CC CC -----
```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AE000925; AAB86148.1; -;
 DR PIR; A69091; A69091.
 DR HSSP; O57816; 1FSZ.
 DR InterPro; IPR000158; FtsZ.
 DR InterPro; IPR003008; Tubulin_FtsZ.
 DR Pfam; PF00091; tubulin_1.
 DR Pfam; PF03953; tubulin_C_1.
 DR PRINTS; PR00423; CELLDVIFTSZ.
 DR TIGRFAMs; TIGR00065; ftsz_1.
 DR PROSITE; PS01134; FTSZ_1; 1.
 DR PROSITE; PS01135; FTSZ_2; 1.
 KW Cell division; Septation; GTP-binding; Complete proteome.
 FT NP_BIND 131 139 GTP (POTENTIAL).
 SQ SEQUENCE 381 AA; 40619 MW; 8E96863D4AF012EE CRC64;

Query Match 35.0%; Score 7; DB 1; Length 381;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GGAGNNT 20
 DB 47 GGAGNNT 53

RESULT 9
 ABFA_STRLI
 ID ABFA_STRLI STANDARD; PRT; 662 AA.
 AC P53627;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Alpha-L-arabinofuranosidase (EC 3.2.1.55) (Arabinosidase) (ABF)
 DE (Alpha-L-AP).
 GN ABFA.
 OS Streptomyces lividans.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1916;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-23.
 RC STRAIN=66 / 1326;
 RX MEDLINE=94379974; PubMed=8092996;
 RA Manin C., Shareek F., Morosoli R., Kluepfel D.,
 RT "Purification and characterization of an alpha-L-arabinofuranosidase
 from Streptomyces lividans 66 and DNA sequence of the gene (abfa).";
 RL Biochem. J. 302:443-449(1994).
 CC -!- FUNCTION: Important role in the degradation of arabinose-
 CC containing hemicelluloses. Hydrolyzes rapidly the short-chain
 CC arabinio-oligosaccharides from digestion of xylan with xylanases.
 CC Acts slowly on arabinan and arabinoxylan from wheat and rye flour
 CC to release L-arabinofuranose. Does not act on oat-spelts xylan or
 CC arabinogalactan.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing alpha-L-
 CC arabinofuranoside residues in alpha-L-arabinosides.
 CC -!- SUBUNIT: Oligomer.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- MISCELLANEOUS: Optimal enzyme activity occurs at 60 degrees
 CC Celsius and pH 6.0.
 CC -!- SIMILARITY: BELONGS TO FAMILY 51 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U04630; AAG61708.1; -;
 DR PIR; S55274; S55274.
 KW Hydrolase; Glycosidase.
 SQ SEQUENCE 662 AA; 72496 MW; DAAF66A577C1D6D1 CRC64;

Query Match 35.0%; Score 7; DB 1; Length 662;
 Best Local Similarity 100.0%; Pred. No. 3.3;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 TKVCGAP 9
 DB 565 TKVCGAP 571

RESULT 10
 HLYI_ECOLI
 ID HLYI_ECOLI STANDARD; PRT; 1023 AA.
 AC P09983;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hemolysin, chromosomal.
 GN HLYA.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J96 / Serotype O4;
 RX MEDLINE=85234404; PubMed=3891743;
 RA Felmler T., Pellett S., Welch R.A.;
 RT "Nucleotide sequence of an Escherichia coli chromosomal hemolysin.";
 RL J. Bacteriol. 163:94-105(1985).
 RN [2]
 RP SEQUENCE OF 1-44 FROM N.A.
 RC STRAIN=2001;
 RX MEDLINE=85258115; PubMed=3894051;
 RA Nicaud J.-M., Mackman N., Gray L., Holland I.B.;
 RT "Characterisation of HlyC and mechanism of activation and secretion
 of haemolysin from E. coli 2001.";
 RL FEBS Lett. 187:339-344(1985).
 CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
 CC cell membranes and cause cell rupture by mechanisms not clearly
 CC defined.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- DOMAIN: The Gly-rich region is probably involved in binding
 CC calcium, which is required for target cell-binding or cytolytic
 CC activity.
 CC -!- DOMAIN: The three transmembrane domains are believed to be
 CC involved in pore formation by the cytotoxin.
 CC -!- PTM: Palmitoylated by hlyC. The toxin only becomes active when
 CC modified.
 CC -!- MISCELLANEOUS: The hemolysin of E.coli is produced predominantly
 CC by strains causing extraintestinal infections, such as those of
 CC the urinary tract.
 CC -!- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M10133; AAA23975.1; -;
 DR EMBL; X02768; CAA26546.1; -;
 DR PIR; A24433; LEECA.
 DR InterPro; IPR001343; Hemlysn_Ca_bind.

```

DR InterPro; IPR003995; RtxA.
DR Pfam; PF00353; hemolysinCbind; 6.
DR Pfam; PF02382; RTX; 1.
DR PRINTS; PR00313; CABNDNGRPT.
DR PRINTS; PR01488; RTXTOXINA.
DR PROSITE; PS00330; HEMOLYSIN_CALCIUM; 4.
KW Hemolysis; Toxin; Cytolysis; Repeat; Calcium; Transmembrane;
KW Lipoprotein; Palmitate.
FT TRANSMEM 237 259
FT TRANSMEM 267 326
FT TRANSMEM 364 410
FT TRANSMEM 723 869
FT DOMAIN 723 728
FT REPEAT 723 728
FT REPEAT 732 737
FT REPEAT 741 746
FT REPEAT 750 755
FT REPEAT 759 764
FT REPEAT 768 773
FT REPEAT 777 782
FT REPEAT 786 791
FT REPEAT 795 800
FT REPEAT 806 812
FT REPEAT 816 821
FT REPEAT 825 830
FT REPEAT 834 839
FT REPEAT 843 848
FT REPEAT 855 860
FT REPEAT 864 869
FT LIPID 563 563
FT LIPID 689 689
FT LIPID 6 6
FT VARIANT 6 6
SQ SEQUENCE 1023 AA; 109867 MW; 196D5C0A9A28B54D CRC64;

Query Match 35.0%; Score 7; DB 1; Length 1023;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 IGGAGNN 19
Db 785 IGGAGNN 791

RESULT 11
PUR6_MYCTU STANDARD; PRT; 174 AA.
AC P96880;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphoribosylaminoimidazole carboxylase catalytic subunit
DE (EC 4.1.1.21) (AIR carboxylase) (AIRCL)
GN PURE OR RV3275C OR MT3375 OR MTCY71.15C.
OS Mycobacterium tuberculosis
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Bv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.

STRAIN=CDC 1551 / Oehkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS SUBUNIT CAN ALONE TRANSFORM AIR TO CAIR, BUT IN
ASSOCIATION WITH PURK, WHICH POSSESSES AN ATPASE ACTIVITY, AN
ENZYME COMPLEX IS PRODUCED WHICH IS CAPABLE OF CONVERTING AIR TO
CAIR EFFICIENTLY UNDER PHYSIOLOGICAL CONDITION (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 5-amino-1-(5-phospho-D-riboseyl)imidazole-4-
carboxylate = 5-amino-1-(5-phospho-D-riboseyl)imidazole + CO(2).
CC -!- PATHWAY: De novo purine biosynthesis; sixth step.
CC -!- SUBUNIT: Homooctamer (By similarity).
CC -!- SIMILARITY: TO OTHER BACTERIAL PURE, ALSO TO AIR CARBOXYLASE FROM
FUNGI.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL; Z92771; CAB07078.1; -.
EMBL; AE007147; AA047716.1; -.
PIR; D70979; D70979.
HSP; P09028; IQCZ.
TIGR; MT3375; -.
TubercuList; RV3275C; -.
InterPro; IPR000031; AIR_carboxyl.
Pfam; PF00731; AIRC; 1.
PRODOM; PD002193; AIR_carboxyl; 1.
TIGRFAMs; TIGR01162; pure; 1.
Purine biosynthesis; Lyase; Decarboxylase; Complete proteome.
SQ SEQUENCE 174 AA; 17674 MW; 18D5F357FA7BC8D1 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 174;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 IGGAGN 18
Db 121 IGGAGN 126

RESULT 12
FTSZ_THEME STANDARD; PRT; 351 AA.
AC O08358;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cell division protein ftsZ.
GN FTSZ OR TW0836.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=98267010; PubMed=9605973;
RA Lu C., Stricker J., Erickson H.P.;
RA "FtsZ from Escherichia coli, Azotobacter vinelandii, and Thermotoga
maritima -- quantitation, GTP hydrolysis, and assembly.";
RL Cell Motil. Cytoskeleton 40:71-86(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;

```

RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M., White O.,
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of *Thermotoga maritima*.";
RL Nature 399:323-329 (1999).
CC -!- FUNCTION: This protein is essential to the cell-division process.
CC It seems to assemble into a dynamic ring on the inner surface of
CC the cytoplasmic membrane at the place where division will occur,
CC and the formation of the ring is the signal for septation to occur.
CC begin. Binds to and hydrolyzes GTP.
CC -!- SUBUNIT: Aggregates to form a ring-like structure.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Assembles at the inner surface
CC of the cytoplasmic membrane (by similarity).
CC -!- SIMILARITY: BELONGS TO THE FTSZ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U65944; AAC24604.1; -.
DR EMBL; AE001750; AAD35918.1; -.
DR PIR; H72328; H72328.
DR HSP; Q57816; IFSZ.
DR TIGR; TM0836; -.
DR InterPro; IPR000158; FtsZ.
DR InterPro; IPR003008; Tubulin_FtsZ.
DR Pfam; PF00091; tubulin; 1.
DR Pfam; PF03953; tubulin_C; 1.
DR PRINTS; PR00423; CELLDVIFTSZ.
DR TIGRfam; TIGR00065; ftsz; 1.
DR PROSITE; PS01134; FTSZ_1; 1.
DR PROSITE; PS01135; FTSZ_2; 1.
KW Cell division; Septation; GTP-binding; Complete proteome.
FT NP_BIND 114 122 GTP (POTENTIAL).
FT CONFLICT 102 102 V -> A (IN REF. 1).
SQ SEQUENCE 351 AA; 38307 MW; 5F2E11FDA92AFE7 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 351;
Best Local Similarity 100.0%; Pred. No. 22; Mismatches 0; Indels 0; Gaps 0;
Matches 6; Conservative 0;
QY 14 GGAGNN 19
DB 30 GGAGNN 35

RESULT 13
ID TBG3 MAIZE STANDARD; PRT; 421 AA.
AC Q41874.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tubulin gamma-3 chain (Gamma-3 tubulin) (Fragment).
GN TUBG3 OR TUBG OR TUBC.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Black Mexican Sweet;
RA Canaday J., Stoppin V., Endle M.C., Lambert A.M.;
RT "Identification of two maize cDNAs encoding gamma-tubulin.";

RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: TUBULIN IS THE MAJOR CONSTITUT OF MICROTUBULES.
CC GAMMA TUBULIN IS FOUND AT MICROTUBULE ORGANIZING CENTERS (MTOC)
CC SUCH AS THE SPINDLE POLES OR THE CENTROSOME, SUGGESTING THAT IT
CC IS INVOLVED IN THE MINUS-END NUCLEATION OF MICROTUBULE ASSEMBLY.
CC -!- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X83696; CAA58671.1; -.
DR InterPro; IPR000217; Tubulin.
DR InterPro; IPR003008; Tubulin_FtsZ.
DR Pfam; PF00091; tubulin; 1.
DR Pfam; PF03953; tubulin_C; 1.
DR PRINTS; PR01161; TUBULIN.
DR PROSITE; PS00227; TUBULIN; 1.
KW Microtubules; GTP-binding; Multigene family.
FT NON_TER 1 1
FT NP_BIND 94 100 GTP (POTENTIAL).
SQ SEQUENCE 421 AA; 47541 MW; 7C2BC3B91EC77523 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 421;
Best Local Similarity 100.0%; Pred. No. 26; Mismatches 0; Indels 0; Gaps 0;
Matches 6; Conservative 0;
QY 14 GGAGNN 19
DB 50 GGAGNN 55

RESULT 14
ID TOLB HAEIN STANDARD; PRT; 427 AA.
AC P44677; P94811;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE TOLB protein precursor.
DE TOLB OR HI0382.
GN Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7942800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1479;
RX MEDLINE=97080550; PubMed=8921895;
RA Sen K., Sikkema D.J., Murphy T.F.;
RT "Isolation and characterization of the Haemophilus influenzae tolQ,
RT tolR, tolA and tolB genes.";
RL Gene 178:75-81 (1996).

CC -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE TOLB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32722; AAC22040.1; -.
DR EMBL; U32470; AAC44597.1; -.
DR PIR; F64064; F64064.
DR HSSP; P19335; 1CRZ.
DR TIGR; H10382; -.
DR PFAM; PF04052; TolB_N; 1.
KW Transport; Protein transport; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 23
FT CHAIN 24 427
FT VARIAT 6 6
FT VARIAT 14 14
FT VARIAT 17 19
FT VARIAT 21 21
FT VARIAT 79 79
FT VARIAT 129 129
FT VARIAT 150 160
FT VARIAT 237 237
FT VARIAT 322 322
FT VARIAT 326 326
FT VARIAT 328 328
SQ SEQUENCE 427 AA; 44967 MW; 0882201AEB9254B9 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 427;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GAGNNT 20
DB 288 GAGNNT 293
|||||

RESULT 15
TBG_CAEEL STANDARD; PRT; 444 AA.
AC P34475;
DT 01-FEB-1994 (Rel. 28, Created).
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Tubulin gamma chain (Gamma tubulin).
GN TBG-1 OR F58A4.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematozoa; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20490776; PubMed=11034903;
RA Bobinnec Y., Fukuda M., Nishida E.;
RT "Identification and characterization of Caenorhabditis elegans
RT gamma-tubulin in dividing cells and differentiated tissues.";
RL J. Cell Sci. 113:3747-3759(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kersey J., Kirsten J., Laibster N.,

RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woulman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -!- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES.
CC GAMMA TUBULIN IS FOUND AT MICROTUBULE ORGANIZING CENTERS (MTOC)
CC SUCH AS THE SPINDLE POLES OR THE CENTROSOME, SUGGESTING THAT IT
CC IS INVOLVED IN THE MINUS-END NUCLEATION OF MICROTUBULE ASSEMBLY.
CC -!- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF287259; AAG24513.1; -.
DR EMBL; Z22179; CA80164.1; -.
DR PIR; S40980; S40980.
DR WormPep; F58A4.8; CE00224.
DR InterPro; IPR000217; Tubulin.
DR Pfam; PF00091; tubulin_1.
DR PRINTS; PR01161; TUBULIN.
DR PROSITE; PS00227; TUBULIN; 1.
KW Microtubules; GTP-binding.
FT NP_BIND 144 150
FT GTP (POTENTIAL).
SQ SEQUENCE 444 AA; 49929 MW; F96CB905B014D6A3 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 444;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 GGAGNN 19
DB 100 GGAGNN 105
|||||

Search completed: November 21, 2003, 20:59:59
Job time : 4.85 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:45:57 ; Search time 24.4 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-84
Perfect score: 20
Sequence: 1 GFTKVCAPPVIGGAGNNT 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	10	50.0	128 12 Q8JYQ3	Q8JYQ3 hepatitis c
2	10	50.0	128 12 Q8JYQ4	Q8JYQ4 hepatitis c
3	10	50.0	129 12 Q8JYQ1	Q8JYQ1 hepatitis c
4	10	50.0	129 12 Q8JYP9	Q8JYP9 hepatitis c
5	10	50.0	129 12 Q8JYQ0	Q8JYQ0 hepatitis c
6	10	50.0	137 12 Q8JYQ6	Q8JYQ6 hepatitis c
7	10	50.0	137 12 Q8J4W1	Q8J4W1 hepatitis c
8	10	50.0	137 12 Q914Q5	Q914Q5 hepatitis c
9	10	50.0	137 12 Q914Q8	Q914Q8 hepatitis c
10	10	50.0	137 12 Q914W4	Q914W4 hepatitis c
11	10	50.0	137 12 Q914W9	Q914W9 hepatitis c
12	10	50.0	137 12 Q914W2	Q914W2 hepatitis c
13	10	50.0	137 12 Q914W8	Q914W8 hepatitis c
14	10	50.0	137 12 Q914Q4	Q914Q4 hepatitis c
15	10	50.0	137 12 Q914W5	Q914W5 hepatitis c
16	10	50.0	137 12 Q914W0	Q914W0 hepatitis c

17	50.0	137 12	Q914Q0	Q914Q0 hepatitis c	
18	10	50.0	137 12	Q914X0	Q914X0 hepatitis c
19	10	50.0	137 12	Q914V9	Q914V9 hepatitis c
20	10	50.0	137 12	Q914V8	Q914V8 hepatitis c
21	10	50.0	137 12	Q914Q3	Q914Q3 hepatitis c
22	10	50.0	137 12	Q914W6	Q914W6 hepatitis c
23	10	50.0	137 12	Q914Q2	Q914Q2 hepatitis c
24	10	50.0	137 12	Q914Q1	Q914Q1 hepatitis c
25	10	50.0	137 12	Q914W7	Q914W7 hepatitis c
26	10	50.0	137 12	Q914Q7	Q914Q7 hepatitis c
27	10	50.0	137 12	Q914V7	Q914V7 hepatitis c
28	10	50.0	137 12	Q914W3	Q914W3 hepatitis c
29	10	50.0	206 12	Q81571	Q81571 hepatitis c
30	10	50.0	403 12	Q9PX22	Q9PX22 hepatitis c
31	10	50.0	420 12	Q98UN4	Q98UN4 hepatitis c
32	10	50.0	640 12	Q68966	Q68966 hepatitis c
33	10	50.0	746 12	Q8JPM2	Q8JPM2 hepatitis c
34	10	50.0	778 12	Q04185	Q04185 hepatitis c
35	10	50.0	778 12	Q04184	Q04184 hepatitis c
36	10	50.0	2436 12	Q81756	Q81756 hepatitis c
37	10	50.0	3011 12	Q91FE5	Q91FE5 hepatitis c
38	10	50.0	3011 12	Q36579	Q36579 hepatitis c
39	10	50.0	3011 12	Q36610	Q36610 hepatitis c
40	10	50.0	3011 12	Q03463	Q03463 hepatitis c
41	10	50.0	3011 12	Q9ELS8	Q9ELS8 hepatitis c
42	10	50.0	3011 12	Q9DIT6	Q9DIT6 hepatitis c
43	10	50.0	3011 12	Q36608	Q36608 hepatitis c
44	8	40.0	128 12	Q8JYQ2	Q8JYQ2 hepatitis c
45	8	40.0	133 12	Q81497	Q81497 hepatitis c

ALIGNMENTS

RESULT 1

Q8JYQ3 PRELIMINARY; PRT; 128 AA.
ID Q8JYQ3
AC Q8JYQ3;
DT 01-OCT-2002 (TrenbLrel. 22, Created)
DT 01-OCT-2002 (TrenbLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrenbLrel. 23, Last annotation update)
DE E2 protein (Genome polyprotein) (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LC2;
RA Quer J., Murillo P., Esteban J.I., Martell M., Esteban R., Guardia J.;
RT "Sexual transmission of hepatitis C virus from a chronic patient to
RT his sexual partner after removal of an intrauterine device."
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF506807; AAM33354.1; -
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 128 128
SQ SEQUENCE 128 AA; 14481 MW; 7BF0E9266B8C9C0F CRC64;

Query Match 50.0%; Score 10; DB 12; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.0062;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
Db 3 GFTKVCAPP 12

RESULT 2

```

QBQYQ4
ID QBQYQ4 PRELIMINARY; PRT; 128 AA.
AC QBQYQ4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE E2 protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LC4;
RA Quer J., Murillo P., Esteban J.I., Martell M., Esteban R., Guardia J.;
RT "Sexual transmission of hepatitis C virus from a chronic patient to
his sexual partner after removal of an intrauterine device.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF506806; AAM33353.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR CoRoT; CR000000; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 128
SQ SEQUENCE 128 AA; 14481 MW; 7EF65920DD3C9ABF CRC64;

Query Match 50.0%; Score 10; DB 12; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.0062;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
Db 3 GFTKVCAPP 12

RESULT 3
QBQYQ1
ID QBQYQ1 PRELIMINARY; PRT; 129 AA.
AC QBQYQ1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE E2 protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LC4;
RA Quer J., Murillo P., Esteban J.I., Martell M., Esteban R., Guardia J.;
RT "Sexual transmission of hepatitis C virus from a chronic patient to
his sexual partner after removal of an intrauterine device.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF506809; AAM33356.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 129
SQ SEQUENCE 129 AA; 14557 MW; CEB9F0F5A16ADB2B CRC64;

Query Match 50.0%; Score 10; DB 12; Length 129;
Best Local Similarity 100.0%; Pred. No. 0.0062;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
Db 3 GFTKVCAPP 12

QBQYQ4
ID QBQYQ4 PRELIMINARY; PRT; 128 AA.
AC QBQYQ4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE E2 protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LC4;
RA Quer J., Murillo P., Esteban J.I., Martell M., Esteban R., Guardia J.;
RT "Sexual transmission of hepatitis C virus from a chronic patient to
his sexual partner after removal of an intrauterine device.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF506806; AAM33353.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 128
SQ SEQUENCE 128 AA; 14481 MW; 7EF65920DD3C9ABF CRC64;

Query Match 50.0%; Score 10; DB 12; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.0062;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
Db 3 GFTKVCAPP 12

QBQYQ1
ID QBQYQ1 PRELIMINARY; PRT; 129 AA.
AC QBQYQ1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE E2 protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LC4;
RA Quer J., Murillo P., Esteban J.I., Martell M., Esteban R., Guardia J.;
RT "Sexual transmission of hepatitis C virus from a chronic patient to
his sexual partner after removal of an intrauterine device.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF506809; AAM33356.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 129
SQ SEQUENCE 129 AA; 14557 MW; CEB9F0F5A16ADB2B CRC64;

Query Match 50.0%; Score 10; DB 12; Length 129;
Best Local Similarity 100.0%; Pred. No. 0.0062;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
Db 3 GFTKVCAPP 12

```

```

RESULT 4
QBQYQ9
ID QBQYQ9 PRELIMINARY; PRT; 129 AA.
AC QBQYQ9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE E2 protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LC4;
RA Quer J., Murillo P., Esteban J.I., Martell M., Esteban R., Guardia J.;
RT "Sexual transmission of hepatitis C virus from a chronic patient to
his sexual partner after removal of an intrauterine device.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF506811; AAM33358.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 129
SQ SEQUENCE 129 AA; 14639 MW; F84BAFF5F0A1C2DF CRC64;

Query Match 50.0%; Score 10; DB 12; Length 129;
Best Local Similarity 100.0%; Pred. No. 0.0062;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
Db 3 GFTKVCAPP 12

RESULT 5
QBQYQ0
ID QBQYQ0 PRELIMINARY; PRT; 129 AA.
AC QBQYQ0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE E2 protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LC4;
RA Quer J., Murillo P., Esteban J.I., Martell M., Esteban R., Guardia J.;
RT "Sexual transmission of hepatitis C virus from a chronic patient to
his sexual partner after removal of an intrauterine device.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF506810; AAM33357.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 129
SQ SEQUENCE 129 AA; 14605 MW; 865657563EC97B7A CRC64;

Query Match 50.0%; Score 10; DB 12; Length 129;
Best Local Similarity 100.0%; Pred. No. 0.0062;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10
Db 3 GFTKVCAPP 12

```

Db 3 GFTKVCAPP 12

RESULT 6

Q914Q6 PRELIMINARY; PRT; 137 AA.
 ID Q914Q6
 AC Q914Q6
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=wt;
 RA Lyra A.C., Fan X., Di Bisceglie A.M.;
 RT "Genetic diversity of two putative cd81 binding regions of HCV before
 and after liver transplantation.";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF431882; AAL30731.1; -;
 DR InterPro: IPR002531; HCV_NSI.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR ProDom: PD186062; HCV_NSI; 1.
 DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 137
 SQ SEQUENCE 137 AA; 15042 MW; 08E5232119C91A09 CRC64;

Query Match 50.0%; Score 10; DB 12; Length 137;
 Best Local Similarity 100.0%; Pred. No. 0.0066;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10

Db 81 GFTKVCAPP 90

RESULT 7

Q914W1 PRELIMINARY; PRT; 137 AA.
 ID Q914W1
 AC Q914W1
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=dg;
 RA Lyra A.C., Fan X., Di Bisceglie A.M.;
 RT "Genetic diversity of two putative cd81 binding regions of HCV before
 and after liver transplantation.";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF431826; AAL30676.1; -;
 DR InterPro: IPR002531; HCV_NSI.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR ProDom: PD186062; HCV_NSI; 1.
 DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 137
 SQ SEQUENCE 137 AA; 14996 MW; 46F1541F17CFF13F CRC64;

Query Match 50.0%; Score 10; DB 12; Length 137;
 Best Local Similarity 100.0%; Pred. No. 0.0066;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10

Db 81 GFTKVCAPP 90

RESULT 8

Q914Q5 PRELIMINARY; PRT; 137 AA.
 ID Q914Q5
 AC Q914Q5
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=wt;
 RA Lyra A.C., Fan X., Di Bisceglie A.M.;
 RT "Genetic diversity of two putative cd81 binding regions of HCV before
 and after liver transplantation.";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF431883; AAL30732.1; -;
 DR InterPro: IPR002531; HCV_NSI.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR ProDom: PD186062; HCV_NSI; 1.
 DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 137
 SQ SEQUENCE 137 AA; 15070 MW; 08F2132F29D01A09 CRC64;

Query Match 50.0%; Score 10; DB 12; Length 137;
 Best Local Similarity 100.0%; Pred. No. 0.0066;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFTKVCAPP 10

Db 81 GFTKVCAPP 90

RESULT 9

Q914Q8 PRELIMINARY; PRT; 137 AA.
 ID Q914Q8
 AC Q914Q8
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=wt;
 RA Lyra A.C., Fan X., Di Bisceglie A.M.;
 RT "Genetic diversity of two putative cd81 binding regions of HCV before
 and after liver transplantation.";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF431880; AAL30729.1; -;
 DR InterPro: IPR002531; HCV_NSI.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR ProDom: PD186062; HCV_NSI; 1.
 DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 137
 SQ SEQUENCE 137 AA; 15070 MW; 08F2132F29D01A09 CRC64;

```

Query Match          50.0%; Score 10; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GFTKVCGAPP 10
Db      81 GFTKVCGAPP 90

RESULT 10
Q914W4
ID Q914W4 PRELIMINARY; PRT; 137 AA.
AC Q914W4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=dg;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431823; AAL30673.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 14989 MW; 0C203D19D57247E4 CRC64;

Query Match          50.0%; Score 10; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GFTKVCGAPP 10
Db      81 GFTKVCGAPP 90

RESULT 11
Q914W9
ID Q914W9 PRELIMINARY; PRT; 137 AA.
AC Q914W9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=dg;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431817; AAL30668.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 14989 MW; 0C203D19D57247E4 CRC64;

Query Match          50.0%; Score 10; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GFTKVCGAPP 10
Db      81 GFTKVCGAPP 90

RESULT 12
Q914W2
ID Q914W2 PRELIMINARY; PRT; 137 AA.
AC Q914W2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=dg;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431825; AAL30675.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 14918 MW; FC232D19902580F9 CRC64;

Query Match          50.0%; Score 10; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GFTKVCGAPP 10
Db      81 GFTKVCGAPP 90

RESULT 13
Q914W8
ID Q914W8 PRELIMINARY; PRT; 137 AA.
AC Q914W8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=dg;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431819; AAL30669.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.

```

```

FT NON_TER 137
SQ SEQUENCE 137 AA; 14961 MW; 0C339F93EE7247E4 CRC64;

Query Match          50.0%; Score 10; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GFTKVCGAPP 10
Db      81 GFTKVCGAPP 90

RESULT 12
Q914W2
ID Q914W2 PRELIMINARY; PRT; 137 AA.
AC Q914W2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=dg;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431825; AAL30675.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 14918 MW; FC232D19902580F9 CRC64;

Query Match          50.0%; Score 10; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GFTKVCGAPP 10
Db      81 GFTKVCGAPP 90

RESULT 13
Q914W8
ID Q914W8 PRELIMINARY; PRT; 137 AA.
AC Q914W8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=dg;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431819; AAL30669.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.

```



```
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 14961 MW; 0C339F93BE7247E4 CRC64;

Query Match      50.0%; Score 10; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFTKVCGAPP 10
Db 81 GFTKVCGAPP 90

RESULT 14
Q914Q4
ID Q914Q4 PRELIMINARY; PRT; 137 AA.
AC Q914Q4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=wt;
RA Lyra A.C.; Fan X.; Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431884; AAL30733.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 15070 MW; 08F2132F29D01A09 CRC64;

Query Match      50.0%; Score 10; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFTKVCGAPP 10
Db 81 GFTKVCGAPP 90

RESULT 15
Q914W5
ID Q914W5 PRELIMINARY; PRT; 137 AA.
AC Q914W5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=dg;
RA Lyra A.C.; Fan X.; Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431822; AAL30672.1; -.

KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 14961 MW; 0C339F93BE7247E4 CRC64;

Query Match      50.0%; Score 10; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFTKVCGAPP 10
Db 81 GFTKVCGAPP 90

Search completed: November 21, 2003, 21:08:20
Job time : 25.45 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:22 ; Search time 30.495 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-85

Perfect score: 19

Sequence: 1 IGGAGNTLHCPTDCKRHP 19

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq 19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	100.0	19	AA018714	Hepatitis C virus
2	15	78.9	20	AA018714	HCV E2 peptide E2-
3	15	78.9	30	AA018714	HCV HepC segment
4	15	78.9	179	AA018529	Protein encoded by
5	15	78.9	192	AA018529	HCV nucleocapsid c
6	15	78.9	254	AA018529	Amino acid sequenc
7	15	78.9	363	AA018529	Amino acid sequenc
8	15	78.9	402	AA018529	Sequence of glycop
9	15	78.9	414	AA018529	HCV CKS-NS1S2 fusi

10	15	78.9	414	22	AA018529	HCV recombinant an
11	15	78.9	454	10	AA018529	Sequence of hepati
12	15	78.9	454	10	AA018529	Sequence encoded b
13	15	78.9	454	21	AA018529	Protein encoded by
14	15	78.9	480	14	AA018529	HCV-1 E2/NS1 prote
15	15	78.9	531	22	AA018529	Chimeric HCV E2661
16	15	78.9	621	14	AA018529	Sequence of subfra
17	15	78.9	622	14	AA018529	HCV CKS-NS1S1-NS1S
18	15	78.9	622	22	AA018529	HCV recombinant an
19	15	78.9	637	24	AA018529	Hepatitis C virus
20	15	78.9	663	17	AA018529	HCV1 E2 + NS2 poly
21	15	78.9	663	20	AA018529	Hepatitis C virus
22	15	78.9	738	14	AA018529	HCV CKS-full lengt
23	15	78.9	738	22	AA018529	HCV recombinant an
24	15	78.9	2010	23	AA018529	HCV HepC cassette
25	15	78.9	2435	13	AA018529	HCV polypeptide 1.
26	15	78.9	2436	10	AA018529	Sequence encoded i
27	15	78.9	2436	10	AA018529	Peptide encoded by
28	15	78.9	2436	13	AA018529	HCV amino acid seq
29	15	78.9	2772	11	AA018529	Hepatitis C virus
30	15	78.9	2772	21	AA018529	Protein encoded by
31	15	78.9	2816	14	AA018529	HCV-1 polypeptide
32	15	78.9	2894	13	AA018529	Composite HCV HC-J
33	15	78.9	2894	16	AA018529	Composite hepatitis
34	15	78.9	2955	11	AA018529	Hepatitis C virus
35	15	78.9	2955	20	AA018529	Amino acid sequenc
36	15	78.9	2955	21	AA018529	Polyprotein encode
37	15	78.9	3011	13	AA018529	Compiled HCV sequ
38	15	78.9	3011	14	AA018529	Hepatitis C virus
39	15	78.9	3011	14	AA018529	HCV genomic amino
40	15	78.9	3011	17	AA018529	Hepatitis C virus
41	15	78.9	3011	18	AA018529	HCV polypeptide
42	15	78.9	3011	19	AA018529	HCV polypeptide
43	15	78.9	3011	23	AA018529	Hepatitis C virus
44	15	78.9	3011	23	AA018529	Hepatitis C virus
45	15	78.9	3011	23	AA018529	HCV polypeptide 1a

ALIGNMENTS

RESULT 1
AA018714
ID AA018714 standard; Peptide; 19 AA.
AC AA018714;
XX
DT 24-OCT-2002 (first entry)
XX
DE Hepatitis C virus E2 protein derived peptide E2-21.
DE Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide;
KW immunostimulant; vaccine.
XX
OS Hepatitis C virus.
XX
PN WO200255548-A2.
XX
PD 18-JUL-2002.
XX
PF 11-JAN-2002; 2002MO-EP00219.
XX
PR 11-JAN-2001; 2001US-260699P.
PR 30-AUG-2001; 2001US-315768P.
XX
XX (INNO-) INNOGENETICS NV.
PA Maertens G, Bosman F, Buyse M;
PI
XX
DR WPI; 2002-599657/64.
XX
PT New therapeutic vaccine compositions comprising at least one purified recombinant hepatitis C virus (HCV) single or specific oligomeric

PT recombinant envelope protein E1 or E2, useful for immunizing humans
 PT from HCV infection -
 PS Example 7; Page 229; 243pp; English.
 CC The present invention relates to new therapeutic vaccine compositions for
 CC inducing hepatitis C virus (HCV)-specific antibodies, comprising a
 CC composition containing at least one purified recombinant HCV single or
 CC specific oligomeric recombinant envelope proteins selected from an E1 and
 CC an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are
 CC useful for inducing HCV-specific antibodies or for immunising humans
 CC against HCV. The recombinant HCV E1 and/or E2 proteins are useful as
 CC vaccines or therapeutics, in HCV screening and confirmatory antibody
 CC tests, for raising antibodies, in the preparation of medicament, and for
 CC in vitro monitoring of HCV disease or prognosing the response to
 CC treatment of patients suffering from HCV infection. The present sequence
 CC is a peptide derived from the proteins of the invention.
 XX
 SQ Sequence 19 AA;

Query Match 100.0%; Score 19; DB 23; Length 19;
 Best Local Similarity 100.0%; Pred. No. 1.7e-12;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDCKHP 19
 |||||
 DB 1 IGGAGNNTLHCPTDCKHP 19

RESULT 2
 AAR91008
 ID AAR91008 standard; peptide; 20 AA.
 AC AAR91008;
 DT 25-SEP-1996 (first entry)
 DE HCV E2 peptide E2-21 for competition studies.
 KW HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;
 KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.
 OS Synthetic.
 XX WO9604385-A2.
 PN 15-FEB-1996.
 PD 31-JUL-1995; 95WO-EP03031.
 XX 29-JUL-1994; 94EP-0870132.
 XX (INNO-) INNOGENETICS NV.
 PA Bosman P, Buyse M, De Martynoff G, Maertens G;
 PI WPI; 1996-129401/13.
 DR
 XX Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope
 PT proteins - in presence of disulphide bond cleavage agent, to
 PT produce proteins suitable for direct use in vaccines or diagnostic
 PT assays of HCV
 XX
 PS Example 7; Page 67; 146pp; English.

CC AAR90965-R90980 and AAR90995-R91015 represent synthetic hepatitis C
 CC virus (HCV) E1 and E2 peptides used in competition studies. This
 CC sequence represents a synthetic E2 peptide, and corresponds to residues
 CC 571-590 of the E2 protein sequence. These sequences are useful for in
 CC vitro monitoring of HCV disease, or prognosis of the response to
 CC interferon treatment of patients suffering from HCV infection. These
 CC sequences compete with the proteins produced by AAT12704-T12709 and
 CC AAT12961-T12974, which are included in vectors for the production of

CC recombinant E1, E2, and E1/E2. The recombinant proteins can then be
 CC isolated and purified by carrying out a disulphide bond cleavage, or a
 CC reduction step with a disulphide bond cleavage agent, after lysis of
 CC recombinant host cells. The constructs containing the purified HCV
 CC envelope proteins can be used for vaccinating humans against HCV, for in
 CC vitro detection of HCV antibodies in a sample, and in a serotyping assay
 CC for detecting one or more serological types of HCV present in a
 CC biological sample. The constructs can also be immobilised on a solid
 CC substrate and incorporated into a reversed phase hybridisation assay for
 CC determining the presence or the genotype of HCV. The new purification
 CC method preserves the conformation of the recombinantly expressed E1, E2
 CC and E1/E2, and eliminates contaminating proteins. Antigens isolated
 CC using this method are more reactive with human sera than those isolated
 CC by known techniques.
 XX
 SQ Sequence 20 AA;

Query Match 78.9%; Score 15; DB 17; Length 20;
 Best Local Similarity 100.0%; Pred. No. 1.8e-08;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15
 |||||
 DB 1 IGGAGNNTLHCPTDC 15

RESULT 3
 AAU84636
 ID AAU84636 standard; Peptide; 30 AA.
 AC AAU84636;
 XX 08-MAY-2002 (first entry)
 DT HCV HepC1a segment 39.
 DE
 XX Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
 KW viral infection; human immunodeficiency virus; melanoma;
 KW bacterial infection; Salmonella; Legionella; parasitic infection;
 XX Trypanosoma; Toxoplasma; Giardia.
 OS Hepatitis C virus.
 XX WO200190197-A1.
 PN 29-NOV-2001.
 PD 25-MAY-2001; 2001WO-AU00622.
 XX 26-MAY-2000; 2000AU-0007761.
 PR (AUSU) UNIV AUSTRALIAN NAT.
 PA Thomson SA, Ramehaw IA;
 PI WPI; 2002-147575/19.
 DR N-PSDB; ABK36474.
 DR
 XX New synthetic polypeptides having several different segments of at
 PT least one parent polypeptide linked together differently compared to
 PT the linkage in the parent polypeptide, for inducing immune response
 PT against a pathogen or cancer -
 XX
 PS Example 2; Fig 26; 364pp; English.

CC The invention relates to a new synthetic polypeptide (I) comprising
 CC several different segments of at least one parent polypeptide linked
 CC together in a different relationship relative to their linkage in the
 CC parent polypeptide to impede, abrogate or otherwise alter at least one
 CC function associated with the parent polypeptide and for inducing an
 CC immune response against a pathogen or cancer. Also included are a
 CC synthetic polynucleotide encoding and a computer system for
 CC designing the synthetic polypeptides. The synthetic polypeptides and

CC polynucleotides are referred to as a Savine. The synthetic polypeptide is
 CC useful for modulating immune responses preferably directed against a
 CC pathogen or a cancer. (e.g., cancers of the lung, breast, ovary, cervix,
 CC colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone
 CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.
 CC Compositions comprising the polypeptide may be used in the treatment or
 CC prophylaxis against viral (such as infections caused by HIV (human
 CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
 CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
 CC (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,
 CC Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic
 CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
 CC Trypanosoma, Toxoplasma and Giardia) infections. The present
 CC sequence is a peptide derived from a parent protein used to
 CC construct a savine of the invention.

XX Sequence 30 AA;

Query Match 78.9%; Score 15; DB 23; Length 30;
 Best Local Similarity 100.0%; Pred. No. 2.5e-08;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15
 |||||
 DB 3 IGGAGNNTLHCPTDC 17

RESULT 4

AAB18529
 ID AAB18529 standard; Protein; 179 AA.

AC AAB18529;

DT 15-JAN-2001 (first entry)

DE Protein encoded by a novel hepatitis C virus cDNA clone 131.

KW Hepatitis C virus; HCV; antisense polynucleotide; polypeptide;
 viral infectivity; viral replication.

OS Hepatitis C virus.

XX EP1034785-A2.

XX 13-SEP-2000.

PF 16-MAR-1990; 2000EP-0109602.

XX 17-MAR-1989; 89US-0325338.

PR 20-APR-1989; 89US-0341334.

PR 18-MAY-1989; 89US-0355002.

XX 16-MAR-1990; 90EP-0302866.

XX (CHIR) CHIRON CORP.

XX Houghton M, Choo Q, Kuo G;

XX WPI: 2000-566891/53.

DR N-PSDB; AAB75285.

XX Novel composition comprising a hepatitis C virus antisense
 PT polynucleotide which is complementary to or corresponds to a sense
 PT strand of the virus genome, and selectively hybridises to it -

PS Example; Fig 5; 75pp; English.

XX The specification describes a pharmaceutical composition which
 CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
 CC HCV is characterized by a positive stranded RNA genome which has
 CC 40% homology at the polypeptide level to a HCV polypeptide. The
 CC antisense polynucleotide binds to cellular polynucleotides which
 CC enhance and/or are required for viral infectivity, replicative
 CC ability or chronicity. The antisense polynucleotides may also be

CC designed to bind with high specificity, to be of increased stability,
 CC to be stable and to have low toxicity. The composition also comprises
 CC an agent which causes viral RNA to be inactive. The composition
 CC is used for preventing HCV replication in a system. The present
 CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
 CC course of the invention.

XX Sequence 179 AA;

Query Match 78.9%; Score 15; DB 21; Length 179;
 Best Local Similarity 100.0%; Pred. No. 1e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15
 |||||
 DB 61 IGGAGNNTLHCPTDC 75

RESULT 5

AAW67009
 ID AAW67009 standard; protein; 192 AA.

XX AAW67009;

DT 02-MAR-1999 (first entry)

DE HCV nucleocapsid core protein.

KW Hepatitis C virus; HCV; nucleocapsid; core protein; analogue; antibody;
 non-structural protein; thioamide bond; peptide bond.

OS Hepatitis C virus.

XX Key Location/Qualifiers

FT Misc-difference 79 /note= "given in specification as Try"

XX JP10226698-A.

XX 25-AUG-1998.

PF 19-FEB-1997; 97JP-0034702.

PR 19-FEB-1997; 97JP-0034702.

XX (KYOW) KYOWA MEDEX KK.

XX WPI: 1998-515103/44.

XX Determination of antibody in sample - uses peptide analog absorbed
 PT or chemically bound on carrier as antigen

XX Disclosure; Page 4; 13pp; Japanese.

XX This sequence represents the Hepatitis C virus (HCV) nucleocapsid core
 CC protein. The invention relates to peptide analogues derived from HCV
 CC proteins, e.g. AAW67417-W67426, which can be used for the determination
 CC of anti-HCV antibodies in a sample. Preferably the peptide analogues
 CC contain one or more thioamide peptide bonds where at least one oxygen
 CC atom of the peptide bond is replaced by sulphur atom. The peptide
 CC analogues can be adsorbed or chemically bound to a carrier.

XX Sequence 192 AA;

Query Match 78.9%; Score 15; DB 19; Length 192;
 Best Local Similarity 100.0%; Pred. No. 1.1e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15
 |||||
 DB 121 IGGAGNNTLHCPTDC 135

```

RESULT 6
AAB68043
ID AAB68043 standard; protein; 254 AA.
XX
XX
AC AAB68043;
XX
XX 29-JUN-2001 (first entry)
XX
XX Amino acid sequence of water soluble variant of envelope E2 protein.
XX
XX Eo protein; HCV; envelope protein; E2 protein; HCV infection;
KW HCV attachment.
XX
XX Synthetic.
OS Hepatitis C virus.
XX WO200122984-A1.
XX
XX 05-APR-2001.
XX
XX 26-SEP-2000; 2000WO-US26395.
XX
XX 29-SEP-1999; 99US-0407430.
XX
XX (UYCO ) UNIV COLUMBIA NEW YORK.
XX Worman HJ, Mamiya N;
XX WPI; 2001-273486/28.
XX
XX Treating or preventing hepatitis C virus infection in a subject,
PT involves administering hepatitis C virus envelope protein E2 binding
PT agents
XX
XX Claim 5; Fig 8; 46pp; English.
XX
XX The present sequence represents a water soluble variant of a Hepatitis C
XX virus (HCV) envelope E2 protein. Agents that bind to the HCV E2 protein
XX (such as the human Eo protein), and so inhibit the attachment of HCV onto
XX cells (especially liver cells), are used to treat HCV infections in
XX mammals, in particular humans. The specification also describes a method
XX for identifying a compound which can be used for treating or preventing
XX HCV in a subject and which can inhibit the attachment of HCV onto cells
XX by inhibiting the binding of HCV envelope E2 protein to a cellular
XX protein associated with HCV attachment and entry into cells. The method
XX comprises incubating the compound, HCV envelope E2 protein or its variant
XX and a cellular protein capable of specifically binding to the HCV E2
XX protein under suitable reaction conditions; determining the interactions
XX between HCV envelope E2 protein and cellular protein in the presence and
XX absence of the compound; and comparing the interaction to identify a
XX compound which can inhibit the attachment of HCV onto cells.
XX
XX SQ Sequence 254 AA;
XX
XX Query Match 78.9%; Score 15; DB 22; Length 254;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-07;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 IGGAGNNTLHCPTDC 15
XX |||||
XX 166 IGGAGNNTLHCPTDC 180
XX
XX
XX RESULT 7
AAB68042
ID AAB68042 standard; protein; 363 AA.
XX
XX AAB68042;
XX
XX 29-JUN-2001 (first entry)
XX
XX Amino acid sequence of a Hepatitis C virus envelope E2 protein.
XX
XX Eo protein; HCV; envelope protein; E2 protein; HCV infection;
KW HCV attachment.
XX
XX Synthetic.
OS Hepatitis C virus.
XX WO200122984-A1.
XX
XX 05-APR-2001.
XX
XX 26-SEP-2000; 2000WO-US26395.
XX
XX 29-SEP-1999; 99US-0407430.
XX
XX (UYCO ) UNIV COLUMBIA NEW YORK.
XX Worman HJ, Mamiya N;
XX WPI; 2001-273486/28.
XX
XX Treating or preventing hepatitis C virus infection in a subject,
PT involves administering hepatitis C virus envelope protein E2 binding
PT agents
XX
XX Claim 5; Fig 8; 46pp; English.
XX
XX The present sequence represents a water soluble variant of a Hepatitis C
XX virus (HCV) envelope E2 protein. Agents that bind to the HCV E2 protein
XX (such as the human Eo protein), and so inhibit the attachment of HCV onto
XX cells (especially liver cells), are used to treat HCV infections in
XX mammals, in particular humans. The specification also describes a method
XX for identifying a compound which can be used for treating or preventing
XX HCV in a subject and which can inhibit the attachment of HCV onto cells
XX by inhibiting the binding of HCV envelope E2 protein to a cellular
XX protein associated with HCV attachment and entry into cells. The method
XX comprises incubating the compound, HCV envelope E2 protein or its variant
XX and a cellular protein capable of specifically binding to the HCV E2
XX protein under suitable reaction conditions; determining the interactions
XX between HCV envelope E2 protein and cellular protein in the presence and
XX absence of the compound; and comparing the interaction to identify a
XX compound which can inhibit the attachment of HCV onto cells.
XX
XX SQ Sequence 254 AA;
XX
XX Query Match 78.9%; Score 15; DB 22; Length 254;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-07;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 IGGAGNNTLHCPTDC 15
XX |||||
XX 166 IGGAGNNTLHCPTDC 180
XX
XX
XX RESULT 8
AAB34438
ID AAB34438 standard; protein; 402 AA.
XX
XX AAB34438;
XX
XX 25-MAR-2003 (updated)
XX 09-AUG-1993 (first entry)
XX
XX Sequence of glycoprotein E2/NS1 in clone HCV1.
XX
XX Hepatitis C virus; envelope protein; glycoprotein; E2/NS1;
KW diagnostic reagent.
XX
XX Hepatitis C virus.
XX
XX EP537626-A1.
XX
XX 21-APR-1993.
XX
XX 08-OCT-1992; 92EP-0117191.
XX
XX
XX Eo protein; HCV; envelope protein; E2 protein; HCV infection;
KW HCV attachment.
XX
XX Hepatitis C virus.
XX WO200122984-A1.
XX
XX 05-APR-2001.
XX
XX 26-SEP-2000; 2000WO-US26395.
XX
XX 29-SEP-1999; 99US-0407430.
XX (UYCO ) UNIV COLUMBIA NEW YORK.
XX Worman HJ, Mamiya N;
XX WPI; 2001-273486/28.
XX
XX Treating or preventing hepatitis C virus infection in a subject,
PT involves administering hepatitis C virus envelope protein E2 binding
PT agents
XX
XX Claim 3; Fig 7; 46pp; English.
XX
XX The present sequence represents a Hepatitis C virus (HCV) envelope
XX E2 protein. Agents that bind to the HCV E2 protein (such as the human
XX Eo protein), and so inhibit the attachment of HCV onto cells
XX (especially liver cells), are used to treat HCV infections in mammals,
XX in particular humans. The specification also describes a method for
XX identifying a compound which can be used for treating or preventing
XX HCV in a subject and which can inhibit the attachment of HCV onto cells
XX by inhibiting the binding of HCV envelope E2 protein to a cellular
XX protein associated with HCV attachment and entry into cells. The method
XX comprises incubating the compound, HCV envelope E2 protein or its variant
XX and a cellular protein capable of specifically binding to the HCV E2
XX protein under suitable reaction conditions; determining the interactions
XX between HCV envelope E2 protein and cellular protein in the presence and
XX absence of the compound; and comparing the interaction to identify a
XX compound which can inhibit the attachment of HCV onto cells.
XX
XX SQ Sequence 363 AA;
XX
XX Query Match 78.9%; Score 15; DB 22; Length 363;
XX Best Local Similarity 100.0%; Pred. No. 1.9e-07;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 IGGAGNNTLHCPTDC 15
XX |||||
XX 188 IGGAGNNTLHCPTDC 202
XX
XX
XX RESULT 9
AAB34438
ID AAB34438 standard; protein; 402 AA.
XX
XX AAB34438;
XX
XX 25-MAR-2003 (updated)
XX 09-AUG-1993 (first entry)
XX
XX Sequence of glycoprotein E2/NS1 in clone HCV1.
XX
XX Hepatitis C virus; envelope protein; glycoprotein; E2/NS1;
KW diagnostic reagent.
XX
XX Hepatitis C virus.
XX
XX EP537626-A1.
XX
XX 21-APR-1993.
XX
XX 08-OCT-1992; 92EP-0117191.
XX
XX

```

```

XX PR 08-OCT-1991; 91JP-0260824.
XX PA (NAHE-) NAT INST OF HEALTH.
XX PI Harada S, Honda Y, Miyamura T, Saito I;
XX DR WPI; 1993-127516/16.
XX DR N-PSDB; AAQ40330.
XX PT Diagnostic reagent for hepatitis C virus - comprises second
XX PT envelope protein or first non-structural protein encoded by HCV
XX PT gene and has sugar chain
XX PS Claim 2; Pages 30-32; 58pp; English.
XX CC Glycoprotein E2/NS1 is derived from the second envelope protein or
XX CC first non-structural protein encoded by the genome of HCV. The
XX CC nucleic acid is extracted from the serum of the patient of hepatitis
XX CC C. The serum is pref. mixed with transfer RNA (tRNA) as a carrier
XX CC of virus RNA. As a technique of cloning cDNA from the nucleic acid,
XX CC it is preferred to use polymerase chain reaction method. In the
XX CC reaction, any commercially available random primers or synthesized
XX CC DNA having a base sequence similar to that of primer AS1 may be used
XX CC as a primer. Representative examples of sense primers include S1.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 402 AA;

Query Match 78.9%; Score 15; DB 14; Length 402;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15
Db 232 IGGAGNNTLHCPTDC 246

RESULT 9
AAR33589
ID AAR33589 standard; protein; 414 AA.
XX AC AAR33589;
XX DT 25-MAR-2003 (updated)
XX DT 05-JUL-1993 (first entry)
XX DE HCV CKS-NS1S2 fusion antigen.
XX KW Hepatitis C Virus; non-A, non-B hepatitis virus; NANBH;
XX KW non-structural protein; CMP-KDO synthetase; CKS fusion protein;
XX KW CTP-CMP-3-deoxy-manno-octulosonate cytidyl transferase;
XX KW immunoassay; pHCV-65.
XX OS Hepatitis C Virus.
XX PN WO9304088-A1.
XX PD 04-MAR-1993.
XX PF 21-AUG-1992; 92WO-US07188.
XX PR 21-AUG-1991; 91US-0748561.
XX PA (ABBO ) ABBOTT LAB.
XX PI Dailey SH, Desai SM, Devare SG;
XX DR WPI; 1993-093941/11.
XX PT Hepatitis C assay using recombinant NS1 region antigens - for
XX PT detecting antibodies and antigen in body fluids from individuals
XX PT exposed to hepatitis C virus

```

```

XX PS Claim 2; Page 40-42; 175pp; English.
XX CC Six oligonucleotides representing amino acids 565-731 of the HCV
XX CC genome were ligated together and cloned as a 501bp EcoRI/BamHI
XX CC fragment into the CKS fusion vector pUO200. The amino acid sequence
XX CC of this antigen is designated pHCV-65 (i.e. AAR33589). The resultant
XX CC fusion protein HCV CKS-NS1S2 consists of 239 amino acids of CKS,
XX CC eight amino acids contributed by linker DNA sequences and 167 amino
XX CC acids from the NS1 region of the HCV genome. The fusion protein is
XX CC used to detect antibodies and antigen in body fluids from
XX CC individuals exposed to HCV.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 414 AA;

Query Match 78.9%; Score 15; DB 14; Length 414;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15
Db 254 IGGAGNNTLHCPTDC 268

RESULT 10
AAB69008
ID AAB69008 standard; Peptide; 414 AA.
XX AC AAB69008;
XX DT 17-APR-2001 (first entry)
XX DE HCV recombinant antigen pHCV-65 amino acid sequence SEQ ID NO:32.
XX KW Hepatitis C virus; HCV; antigen; detection; antibody.
XX OS Hepatitis C virus.
XX PN US6172189-B1.
XX PD 09-JAN-2001.
XX PF 02-JUN-1997; 97US-0867611.
XX PR 19-NOV-1992; 92US-0989843.
XX PR 10-JAN-1994; 94US-0179896.
XX PR 01-MAY-1996; 96US-0646757.
XX PR 24-AUG-1990; 90US-0572822.
XX PR 07-NOV-1990; 90US-0614069.
XX PR 21-AUG-1991; 91US-0748561.
XX PR 21-AUG-1991; 91US-0748566.
XX PR 29-OCT-1991; 91US-0748565.
XX PA (ABBO ) ABBOTT LAB.
XX PI Devare SG, Desai SM, Casey JM, Dailey SH, Dawson GJ, Gutierrez RA;
XX PI Leeniewski RR, Stewart JL, Rupprecht KR;
XX XX WPI; 2001-122352/13.
XX PT New recombinant antigens representing distinct antigenic regions of
XX PT Hepatitis C virus (HCV) genome, useful for detection of antibodies and
XX PT antigens in body fluids of individuals exposed to HCV -
XX PS Example 11; Column 141-144; 167pp; English.
XX CC The present invention describes recombinant Hepatitis C virus (HCV)
XX CC antigens (I). (I) is useful as a reagent for the detection of antibodies
XX CC and antigen in body fluids from individuals exposed to HCV. The HCV
XX CC assay uses reliable and efficient reagents and methods to accurately
XX CC detect the presence of HCV antibodies in samples obtained from
XX CC individuals suspected of having HCV infection. AAF32218 to AAF32235,

```

CC AAB51371 to AAB51379 and AAB69001 to AAB69032 represent sequences used
CC in the exemplification of the present invention.

SQ Sequence 414 AA;
Query Match 78.9%; Score 15; DB 22; Length 414;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGAGNNTLHCPTDC 15
| | | | | | | | | | | | | | | | | |
Db 254 IGGAGNNTLHCPTDC 268

RESULT 11
AAP90183
ID AAP90183 standard; protein; 454 AA.

AC AAP90183;
XX
XX
DT 25-MAR-2003 (updated)
DT 01-NOV-1989 (first entry)
XX
XX Sequence of hepatitis C virus cDNA insert in clone k9-1.
DE
XX
XX Hepatitis C virus; clone k9-1; probe; vaccine.
KW
XX Pan troglodytes.

OS
XX
XX
FH Key Location/Qualifiers
FT Region 97..454
XX
XX GB2212511-A.

PN
XX
XX 26-JUL-1989.
PD
XX
XX 18-NOV-1988; 88GB-0027024.

XX
XX 18-NOV-1987; 87US-0122714.
PR 30-DEC-1987; 87US-0139886.
PR 26-FEB-1988; 88US-0161072.
PR 26-OCT-1988; 88US-0263584.

XX (CHIR) CHIRON CORPORATION.

XX Houghton M, Choo QL, Kuo G;

XX WPI; 1989-215054/30.

DR N-PSDB; AAN90335.

XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT polypeptide(s) and antibodies for diagnosis, prevention and
PT treatment of infection.

XX Disclosure; fig 46; 235pp; English.

XX The sequence is the peptide encoded by the hepatitis C virus
CC (HCV) cDNA insert in clone k9-1 (see AAN90335). The polypeptides
CC are used to diagnose HCV-induced NANBH, to raise antibodies for
CC immunoassay or treatment, or to produce vaccines.
CC The region shown overlaps the cDNA of AAN90327.
CC (Updated on 25-MAR-2003 to correct PR field.)

SQ Sequence 454 AA;

Query Match 78.9%; Score 15; DB 10; Length 454;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGAGNNTLHCPTDC 15
| | | | | | | | | | | | | | | | | |
Db 121 IGGAGNNTLHCPTDC 135

RESULT 12

ID AAP92049 standard; protein; 454 AA.

XX AAP92049;

XX 25-MAR-2003 (updated)

DT 21-JAN-1991 (first entry)

XX Sequence encoded by segment of the hepatitis C virus (HCV) cDNA sequence
DE in clone K9-1.

XX Non-a non-B hepatitis; probe; vaccine; diagnosis;
KW passive immunotherapy; antigen.

XX Hepatitis C virus.

XX EP318216-A.

XX 31-MAY-1989.

XX 18-NOV-1988; 88EP-0310922.

XX 18-NOV-1987; 87US-0122714.

PR 30-DEC-1987; 87US-0139886.

PR 26-FEB-1988; 88US-0161072.

PR 06-MAY-1988; 88US-0191263.

PR 26-OCT-1988; 88US-0263584.

PR 14-NOV-1988; 88US-0271450.

XX (CHIR) CHIRON CORP.

XX Houghton M, Choo QL, Kuo G;

XX WPI; 1989-159274/22.

DR N-PSDB; AAN92105.

XX Purified hepatitis C virus -

PT and assorted nucleic acids and polypeptide(s)

PS Example; Fig 46-1 - 46-2; 139pp; English.

XX Purified hepatitis C virus (HCV) and purified or recombinant HCV nucleic
CC acids (NAs), encoding HCV polynucleotides or epitopes, and polypeptides
CC are claimed. HCV is a causative agent of non-A, non-B hepatitis (NANBH).
CC The NAs may be used to design probes for detn. of HCV NAs in samples.
CC The polypeptides may be used as immunoassay reagents and vaccines, and
CC to produce antibodies useful for diagnosis and passive immunotherapy.
CC The purified virus may also be used in vaccines.

CC (Updated on 25-MAR-2003 to correct PR field.)

CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 454 AA;

Query Match 78.9%; Score 15; DB 10; Length 454;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGAGNNTLHCPTDC 15
| | | | | | | | | | | | | | | | | |
Db 121 IGGAGNNTLHCPTDC 135

RESULT 13

AAB18526

ID AAB18526 standard; Protein; 454 AA.

XX AAB18526;

XX 15-JAN-2001 (first entry)

XX Protein encoded by a novel hepatitis C virus cDNA clone k9-1.

XX Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
 KW viral infectivity; viral replication.

OS Hepatitis C virus.

XX EP1034785-A2.

PN 13-SEP-2000.

XX 16-MAR-1990; 2000EP-0109602.

XX 17-MAR-1989; 89US-0325338.

PR 20-APR-1989; 89US-0341334.

PR 18-MAY-1989; 89US-0355002.

PR 16-MAR-1990; 90EP-0302866.

XX (CHIR) CHIRON CORP.

XX Houghton M, Choo Q, Kuo G;

XX WPI; 2000-566891/53.

DR N-PSDB; AAA75282.

XX Novel composition comprising a hepatitis C virus antisense

PT polynucleotide which is complementary to or corresponds to a sense

PT strand of the virus genome, and selectively hybridises to it -

XX Example; Fig 2; 75pp; English.

XX The specification describes a pharmaceutical composition which
 CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
 CC HCV is characterized by a positive stranded RNA genome which has
 CC 40% homology at the polypeptide level to a HCV polyprotein. The
 CC antisense polynucleotide binds to cellular polynucleotides which
 CC enhance and/or are required for viral infectivity, replicative
 CC ability or chronicity. The antisense polynucleotides may also be
 CC designed to bind with high specificity, to be of increased stability,
 CC to be stable and to have low toxicity. The composition also comprises
 CC an agent which causes viral RNA to be inactive. The composition
 CC is used for preventing HCV replication in a system. The present
 CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
 CC course of the invention.

XX Sequence 454 AA;

Query Match 78.9%; Score 15; DB 21; Length 454;

Best Local Similarity 100.0%; Pred. No. 2.1e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15

Db 121 IGGAGNNTLHCPTDC 135

RESULT 14

AAE02622

ID AAR33992 standard; Protein; 480 AA.

XX AAR33992;

AC 25-MAR-2003 (updated)

DT 26-JUL-1993 (first entry)

XX HCV-1 E2/NS1 protein.

XX Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
 KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
 KW domain; immunological; cross-reactive; envelope protein; vaccine;
 KW gp53 (BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.

XX Synthetic.

XX

PN WO9306126-A1.

XX 01-APR-1993.

XX 11-SEP-1992; 92WO-US07683.

XX 13-SEP-1991; 91US-0759575.

XX (CHIR) CHIRON CORP.

XX Houghton M, Weiner AJ;

XX WPI; 1993-117468/14.

XX Immuno-reactive hepatitis C virus polypeptide compans. - contg.

PT at least 2 sequences from the first variable domain of distinct

PT HCV isolates

XX Disclosure; Fig 3; 106pp; English.

XX The sequences given in AAR33992-002 represent a portion of the E2/NS1
 CC protein encoded by group I and group II HCV isolates, from amino acid
 CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
 CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
 CC 30 amino acids which shows large variation between nearly all isolates.
 CC This is an important immunoreactive domain. This putative envelope
 CC glycoprotein E2/NS1 may correspond to the gp53 (BVDV)/gp55 (hog cholera
 CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
 CC flaviviruses, both of which confer protective immunity in hosts
 CC vaccinated with these polypeptides. It has been discovered that a
 CC number of important HCV epitopes vary among viral isolates and that
 CC these epitopes can be mapped to specific domains. This meant that
 CC immunologically cross-reactive polypeptides which focus on variable
 CC rather than constant domains can be produced. See also AAR339134-48
 CC and AAR33982-91.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 480 AA;

Query Match 78.9%; Score 15; DB 14; Length 480;

Best Local Similarity 100.0%; Pred. No. 2.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15

Db 202 IGGAGNNTLHCPTDC 216

RESULT 15

AAE02622

ID AAE02622 standard; Protein; 531 AA.

XX AAE02622;

XX 06-AUG-2001 (first entry)

XX Chimeric HCV E2661-HBsAg S domain encoded by pCMV-II-E2661-SAg.

XX virus-like particle; immunogen; hepatitis B virus surface antigen;

XX HBsAg; HBV; chimeric antigen; hepatitis C virus; HCV; virucide; vaccine;

XX plasmid pCMV-II-E2661-SAg; E2 envelope glycoprotein; S domain.

XX Chimeric - Hepatitis C virus.

XX Chimeric - Hepatitis B virus.

XX Key Location/Qualifiers

FT Region 1..303

FT Region /note= "HCV 661 E2 envelope glycoprotein"

FT Region 306..531

FT Region /note= "HBsAg S domain"

XX WO200138358-A2.

PD 31-MAY-2001.
XX
PF 22-NOV-2000; 2000WO-US32249.
XX
PR 24-NOV-1999; 99US-0167224.
XX
PA (CHIR) CHIRON CORP.
XX
PI Selby M, Glazer E, Houghton M;
XX
DR WPI: 2001-367661/38.
DR N-PSDB; AAD06793.
XX
PT Virus-like particle for use as an immunogen, comprising a first
PT hepatitis B virus surface antigen (HBsAg) and chimeric antigen
PT comprising a second HBsAg covalently linked to hepatitis C immunogenic
PT polypeptide -
XX
PS Claim 29; Fig 4; 115pp; English.
XX
CC The invention relates to a virus-like particle for use as an immunogen,
CC comprising a first hepatitis B virus surface antigen (HBsAg) and a
CC chimeric antigen comprising a second HBsAg which is covalently linked to
CC an hepatitis C virus (HCV) immunogenic polypeptide, where the first and
CC the second HBsAg each comprise a substantially complete S domain.
CC The virus-like particle is useful as immunogen and as vaccine.
CC The present sequence is a chimeric antigen comprising HCV 661 E2
CC envelope glycoprotein and HBsAg S domain encoded by plasmid
CC pCMV-II-E2661-SAg.
XX
SQ Sequence 531 AA;

Query Match 78.9%; Score 15; DB 22; Length 531;
Best Local Similarity 100.0%; Pred. NO. 2.4e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ICGAGNNTLHCPTDC 15
|||
DB 213 ICGAGNNTLHCPTDC 227

Search completed: November 21, 2003, 20:58:06
Job time : 31.545 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.07 Seconds
(without alignments)
79.832 Million cell updates/sec

Title: US-09-973-025-85

Perfect score: 19
Sequence: 1 IGGAGNNTLHCPTDCRKHP 19

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0
Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	100.0	19	3	US-08-612-973-85
2	19	100.0	19	3	US-08-927-597-85
3	15	78.9	20	4	US-08-635-886C-41
4	15	78.9	68	4	US-08-635-886C-250
5	15	78.9	88	1	US-08-440-103-26
6	15	78.9	88	1	US-08-440-103-27
7	15	78.9	88	1	US-08-440-542-26
8	15	78.9	88	1	US-08-440-542-27
9	15	78.9	88	1	US-08-231-368-26
10	15	78.9	88	1	US-08-231-368-27
11	15	78.9	88	1	US-08-440-210-26
12	15	78.9	88	1	US-08-440-210-27
13	15	78.9	88	4	US-09-046-604-26
14	15	78.9	88	4	US-09-046-604-27
15	15	78.9	179	3	US-08-444-818-77
16	15	78.9	402	1	US-08-460-806-13
17	15	78.9	402	1	US-08-325-630-13
18	15	78.9	403	2	US-08-483-695-39
19	15	78.9	403	2	US-07-965-285-39
20	15	78.9	403	2	US-08-487-231-39
21	15	78.9	403	3	US-09-201-912-39
22	15	78.9	414	1	US-07-748-292-8
23	15	78.9	414	3	US-08-867-611-32
24	15	78.9	414	5	PCT-US92-06965A-2
25	15	78.9	454	3	US-08-444-818-73
26	15	78.9	480	1	US-08-440-103-14
27	15	78.9	480	1	US-08-440-542-14

28	15	78.9	480	1	US-08-231-368-14	Sequence 14, Appl
29	15	78.9	480	1	US-08-440-210-14	Sequence 14, Appl
30	15	78.9	480	1	US-09-046-604-14	Sequence 14, Appl
31	15	78.9	621	1	US-07-748-292-7	Sequence 7, Appl
32	15	78.9	622	3	US-08-867-611-34	Sequence 34, Appl
33	15	78.9	622	5	PCT-US92-06965A-4	Sequence 4, Appl
34	15	78.9	663	3	US-08-824-057-3	Sequence 3, Appl
35	15	78.9	663	4	US-09-415-582-3	Sequence 3, Appl
36	15	78.9	663	4	US-09-693-596-4	Sequence 4, Appl
37	15	78.9	738	3	US-08-867-611-35	Sequence 35, Appl
38	15	78.9	738	5	PCT-US92-06965A-5	Sequence 5, Appl
39	15	78.9	2436	3	US-08-444-818-75	Sequence 75, Appl
40	15	78.9	2772	3	US-08-444-818-89	Sequence 89, Appl
41	15	78.9	2894	2	US-08-466-975A-23	Sequence 23, Appl
42	15	78.9	2894	2	US-08-391-671A-23	Sequence 23, Appl
43	15	78.9	2894	3	US-08-467-902A-23	Sequence 23, Appl
44	15	78.9	2894	3	US-09-275-265-23	Sequence 23, Appl
45	15	78.9	2894	4	US-09-941-611-23	Sequence 23, Appl

ALIGNMENTS

RESULT 1
US-08-612-973-85
; Sequence 85, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-612-973-85

Query Match 100.0%; Score 19; DB 3; Length 19;
Best Local Similarity 100.0%; Pred. No. 7.3e-13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 01 IGGAGNNTLHCPTDCRKHP 19
|||||

```
Db      1  IGGAGNNTLHCPTDCRKH 19

RESULT 2
US-08-927-597-85
; Sequence 85, Application US/08927597
; Patent No. 6245503
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-927-597-85

Query Match      100.0%; Score 19; DB 3; Length 19;
Best Local Similarity 100.0%; Pred. No. 7.3e-13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  IGGAGNNTLHCPTDCRKH 19
Db      1  IGGAGNNTLHCPTDCRKH 19

RESULT 3
US-08-635-886C-41
; Sequence 41, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555

Query Match      100.0%; Score 15; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.2e-09;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  IGGAGNNTLHCPTDC 15
Db      1  IGGAGNNTLHCPTDC 15

Query Match      78.9%; Score 15; DB 4; Length 68;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  IGGAGNNTLHCPTDC 15
Db      1  IGGAGNNTLHCPTDC 15

RESULT 5
US-08-440-103-26
; Sequence 26, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,103
FILING DATE: 12-MAY-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-103-26

Query Match 78.9%; Score 15; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15
Db 18 IGGAGNNTLHCPTDC 32

RESULT 6
US-08-440-103-27
Sequence 27, Application US/08440103
Patent No. 5670152
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,103
FILING DATE: 12-MAY-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542

INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-103-27

Query Match 78.9%; Score 15; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15
Db 18 IGGAGNNTLHCPTDC 32

RESULT 7
US-08-440-542-26
Sequence 26, Application US/08440542
Patent No. 5670153
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,542
FILING DATE: 12-MAY-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-542-26

Query Match 78.9%; Score 15; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15
Db 18 IGGAGNNTLHCPTDC 32

RESULT 8
US-08-440-542-27
; Sequence 27, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,542
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-542-27
Query Match 78.9%; Score 15; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 IGGAGNNTLHCPTDC 15
Db 18 IGGAGNNTLHCPTDC 32
RESULT 9
US-08-231-368-26
; Sequence 26, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-231-368-26
Query Match 78.9%; Score 15; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 IGGAGNNTLHCPTDC 15
Db 18 IGGAGNNTLHCPTDC 32
RESULT 10
US-08-231-368-27
; Sequence 27, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids

RESULT 12

;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/046,604
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION NUMBER: US/08/231,368
;; FILING DATE:
;; APPLICATION NUMBER: US 07/759,575
;; FILING DATE: 13-SEP-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: McClung, Barbara G.
;; REGISTRATION NUMBER: 33,113
;; REFERENCE/DOCKET NUMBER: 0205.001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (510) 601-2708
;; TELEFAX: (510) 655-3542
;; INFORMATION FOR SEQ ID NO: 26:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 88 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;;
US-09-046-604-26

Query Match 78.9%; Score 15; DB 4; Length 88;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGAGNNTLHCPTDC 15
Db 18 IGGAGNNTLHCPTDC 32

RESULT 14
US-09-046-604-27
; Sequence 27, Application US/09046604
; Patent No. 6303292
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/046,604
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 655-3542
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:

;;
;; LENGTH: 88 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;;
US-09-046-604-27

Query Match 78.9%; Score 15; DB 4; Length 88;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGAGNNTLHCPTDC 15
Db 18 IGGAGNNTLHCPTDC 32

RESULT 15
US-08-444-818-77
; Sequence 77, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 359-3876
; TELEFAX: (508) 359-3885
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 179 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-444-818-77

Query Match 78.9%; Score 15; DB 3; Length 179;
Best Local Similarity 100.0%; Pred. No. 4.4e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGAGNNTLHCPTDC 15
Db 61 IGGAGNNTLHCPTDC 75

Search completed: November 21, 2003, 21:15:17
Job time : 11.07 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time 18.525 Seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-85

Perfect score: 19
Sequence: 1 IGGAGNNTLHCFTDCRKH 19

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	100.0	19	10	US-09-973-025-85
2	19	100.0	19	11	US-09-973-025-85
3	19	100.0	19	11	US-09-995-808-85
4	19	100.0	19	11	US-09-995-808-85
5	19	100.0	19	12	US-09-995-791-85
6	15	78.9	254	10	US-09-407-430-3
7	15	78.9	363	10	US-09-407-430-2
8	15	78.9	637	12	US-10-187-257-4
9	15	78.9	637	12	US-10-265-083-2
10	15	78.9	2894	10	US-09-941-611-23
11	15	78.9	2894	15	US-10-044-995-23
12	15	78.9	3011	9	US-09-916-359-2
13	15	78.9	3011	16	US-10-232-643-6
14	9	47.4	30	16	US-10-318-200-15
15	9	47.4	290	12	US-10-128-587A-3

16	9	47.4	290	15	US-10-128-590-3	Sequence 3, Appli
17	9	47.4	301	12	US-10-128-587A-5	Sequence 5, Appli
18	9	47.4	301	15	US-10-128-590-5	Sequence 5, Appli
19	9	47.4	314	10	US-09-973-025-42	Sequence 42, Appl
20	9	47.4	314	11	US-09-899-303-42	Sequence 42, Appl
21	9	47.4	314	11	US-09-995-808-42	Sequence 42, Appl
22	9	47.4	314	11	US-09-995-860-42	Sequence 42, Appl
23	9	47.4	314	12	US-09-995-791-42	Sequence 44, Appl
24	9	47.4	319	10	US-09-973-025-44	Sequence 44, Appl
25	9	47.4	319	11	US-09-899-303-44	Sequence 44, Appl
26	9	47.4	319	11	US-09-995-808-44	Sequence 44, Appl
27	9	47.4	319	11	US-09-995-860-44	Sequence 44, Appl
28	9	47.4	319	12	US-09-995-791-44	Sequence 44, Appl
29	9	47.4	338	11	US-09-973-025-38	Sequence 38, Appl
30	9	47.4	338	11	US-09-899-303-38	Sequence 38, Appl
31	9	47.4	338	11	US-09-995-808-38	Sequence 38, Appl
32	9	47.4	338	11	US-09-995-860-38	Sequence 38, Appl
33	9	47.4	338	12	US-09-995-791-38	Sequence 38, Appl
34	9	47.4	343	10	US-09-973-025-40	Sequence 40, Appl
35	9	47.4	343	11	US-09-899-303-40	Sequence 40, Appl
36	9	47.4	343	11	US-09-995-808-40	Sequence 40, Appl
37	9	47.4	343	11	US-09-995-860-40	Sequence 40, Appl
38	9	47.4	343	12	US-09-995-791-40	Sequence 40, Appl
39	9	47.4	347	11	US-09-134-949-9	Sequence 9, Appli
40	9	47.4	363	12	US-10-128-587A-98	Sequence 98, Appl
41	0	47.4	363	15	US-10-128-590-98	Sequence 98, Appl
42	9	47.4	463	10	US-09-973-025-46	Sequence 46, Appl
43	9	47.4	463	11	US-09-899-303-46	Sequence 46, Appl
44	9	47.4	463	11	US-09-995-808-46	Sequence 46, Appl
45	9	47.4	463	11	US-09-995-860-46	Sequence 46, Appl

ALIGNMENTS

RESULT 1
US-09-973-025-85
; Sequence 85, Application US/09973025
; Publication No. US20020182706A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRJOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100

```
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 19 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 85:
US-09-973-025-85

Query Match          100.0%; Score 19; DB 10; Length 19;
Best Local Similarity 100.0%; Pred. No. 4e-13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDCRKHP 19
| | | | | | | | | | | | | | | | |
DB 1 IGGAGNNTLHCPTDCRKHP 19

RESULT 2
US-09-899-303-85
; Sequence 85, Application US/09899303
; Publication No. US20030036110A1
; GENERAL INFORMATION:
;   APPLICANT: MAERTENS, GEERT
;   BOSMAN, FONS
;   DE MARTYNOFF, GUY
;   BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
;   PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: NIXON & VANDERHYE P.C.
;   STREET: 1100 NORTH GLEBE ROAD
;   CITY: ARLINGTON
;   STATE: VIRGINIA
;   COUNTRY: U.S.A.
;   ZIP: 22201-4714
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/899,303
;   FILING DATE: 06-Jul-2001
;   CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/612,973
;   FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
;   NAME: BYRNE, THOMAS E.
;   REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703) 816-4000
;   TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 19 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 85:
US-09-899-303-85

Query Match          100.0%; Score 19; DB 11; Length 19;
Best Local Similarity 100.0%; Pred. No. 4e-13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDCRKHP 19
| | | | | | | | | | | | | | | | |
DB 1 IGGAGNNTLHCPTDCRKHP 19

INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 19 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 85:
US-09-973-025-85

Query Match          100.0%; Score 19; DB 10; Length 19;
Best Local Similarity 100.0%; Pred. No. 4e-13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDCRKHP 19
| | | | | | | | | | | | | | | | |
DB 1 IGGAGNNTLHCPTDCRKHP 19

RESULT 3
US-09-995-808-85
; Sequence 85, Application US/09995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:
;   APPLICANT: Innogenetics N.V.
;   TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
;   therapeutic use.
;   FILE REFERENCE: 2551-70
;   CURRENT APPLICATION NUMBER: US/09/995,808
;   CURRENT FILING DATE: 2001-11-29
;   NUMBER OF SEQ ID NOS: 122
;   SOFTWARE: PatentIn 3.1
;   SEQ ID NO 85
;   LENGTH: 19
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-09-995-808-85

Query Match          100.0%; Score 19; DB 11; Length 19;
Best Local Similarity 100.0%; Pred. No. 4e-13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDCRKHP 19
| | | | | | | | | | | | | | | | |
DB 1 IGGAGNNTLHCPTDCRKHP 19

RESULT 4
US-09-995-860-85
; Sequence 85, Application US/09995860
; Publication No. US20030118603A1
; GENERAL INFORMATION:
;   APPLICANT: Innogenetics N.V.
;   TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
;   therapeutic use.
;   FILE REFERENCE: 2551-69
;   CURRENT APPLICATION NUMBER: US/09/995,860
;   CURRENT FILING DATE: 2001-11-29
;   NUMBER OF SEQ ID NOS: 122
;   SOFTWARE: PatentIn 3.1
;   SEQ ID NO 85
;   LENGTH: 19
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-09-995-860-85

Query Match          100.0%; Score 19; DB 11; Length 19;
Best Local Similarity 100.0%; Pred. No. 4e-13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDCRKHP 19
| | | | | | | | | | | | | | | | |
DB 1 IGGAGNNTLHCPTDCRKHP 19

RESULT 5
US-09-995-791-85
; Sequence 85, Application US/09995791
; Publication No. US20030147918A1
; GENERAL INFORMATION:
;   APPLICANT: Innogenetics N.V.
;   TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
;   therapeutic use.
;   FILE REFERENCE: 2551-88
;   CURRENT APPLICATION NUMBER: US/09/995,791
;   CURRENT FILING DATE: 2001-11-29
;   NUMBER OF SEQ ID NOS: 122
;   SOFTWARE: PatentIn 3.1
;   SEQ ID NO 85
```

```
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-85

Query Match
Best Local Similarity 100.0%; Score 19; DB 12; Length 19;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDCRKH 19
    |||||
Db 1 IGGAGNNTLHCPTDCRKH 19
    |||||

RESULT 6
US-09-407-430-3
; Sequence 3, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-407-430-3

Query Match
Best Local Similarity 100.0%; Score 15; DB 10; Length 254;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15
    |||||
Db 166 IGGAGNNTLHCPTDC 180
    |||||

RESULT 7
US-09-407-430-2
; Sequence 2, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-09-407-430-2

Query Match
Best Local Similarity 100.0%; Score 15; DB 10; Length 363;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15
    |||||
Db 188 IGGAGNNTLHCPTDC 202
    |||||

RESULT 8

US-10-187-257-4
; Sequence 4, Application US/10187257
; Publication No. US20030138458A1
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: O'HAGAN, Derek
; TITLE OF INVENTION: HCV E1E2 VACCINE COMPOSITIONS
; FILE REFERENCE: 2302-17206
; CURRENT APPLICATION NUMBER: US/10/187,257
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
US-10-187-257-4

Query Match
Best Local Similarity 100.0%; Score 15; DB 12; Length 637;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15
    |||||
Db 399 IGGAGNNTLHCPTDC 413
    |||||

RESULT 9
US-10-265-083-2
; Sequence 2, Application US/10265083
; Publication No. US20030170273A1
; GENERAL INFORMATION:
; APPLICANT: O'HAGAN, Derek
; APPLICANT: VALIANTE, Nicholas
; TITLE OF INVENTION: ADJUVANT COMPOSITIONS
; FILE REFERENCE: 2302-18357.30
; CURRENT APPLICATION NUMBER: US/10/265,083
; CURRENT FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7
US-10-265-083-2

Query Match
Best Local Similarity 100.0%; Score 15; DB 12; Length 637;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15
    |||||
Db 399 IGGAGNNTLHCPTDC 413
    |||||

RESULT 10
US-09-941-611-23
; Sequence 23, Application US/09941611
; Patent No. US20020106640A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; POLLET, DIRK
; MAERTENS, GEERT
; VAN HEUVERSUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; ANTIBODIES TO HEPATITIS C VIRUS
```

NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/941,611
FILING DATE: 30-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: 1995-02-21
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-941-611-23

Query Match 78.9%; Score 15; DB 10; Length 2894;
Best Local Similarity 100.0%; Pred. No. 3.4e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15
DB 571 IGGAGNNTLHCPTDC 585
|||||

RESULT 11
US-10-044-995-23
Sequence 23, Application US/10044995
Publication No. US20030049685A1
GENERAL INFORMATION:
APPLICANT: DELEYS, ROBERT J
POLLET, DIRK
MAERTENS, GEERT
VAN HEUVERSWUN, HUGO
TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
ANTIBODIES TO HEPATITIS C VIRUS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/044,995
FILING DATE: 15-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: <Unknown>
APPLICATION NUMBER: US 07/920,286
FILING DATE: 14-OCT-1992
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-044-995-23

Query Match 78.9%; Score 15; DB 15; Length 2894;
Best Local Similarity 100.0%; Pred. No. 3.4e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15
DB 571 IGGAGNNTLHCPTDC 585
|||||

RESULT 12
US-09-916-359-2
Sequence 2, Application US/09916359
Patent No. US20020034734A1
GENERAL INFORMATION:
APPLICANT: Veronique Barban
TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
TREATING C HEPATITIS
FILE REFERENCE: PMCF97-03A
CURRENT APPLICATION NUMBER: US/09/916,359
CURRENT FILING DATE: 2001-07-26
PRIOR APPLICATION NUMBER: 09/386,874
PRIOR FILING DATE: 1999-09-02
PRIOR APPLICATION NUMBER: 97/02,887
PRIOR FILING DATE: 1997-03-06
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 3011
TYPE: PRT
ORGANISM: Virus
US-09-916-359-2

Query Match 78.9%; Score 15; DB 9; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.5e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15
|||||

Db 571 IGGAGNTHLCPTDC 585

RESULT 13
US-10-232-643-6
Sequence 6, Application US/10232643
Publication NO. US20030129586A1
GENERAL INFORMATION:
APPLICANT: HOUGHTON, MICHAEL
CHOO, QUI-LIM
HAN, JANG
CHOE, JOONHO

TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
HELICASE ACTIVITY AND IMPROVED SOLUBILITY

NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSER: CHIRON CORPORATION
STREET: Intellectual Property - R440, P.O. Box 8097
CITY: Emeryville
STATE: California
COUNTRY: U.S.A.
ZIP: 94662-8097

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/10/232,643
FILING DATE: 30-Aug-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/483,799
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/529,169
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0100.005

TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 923-3274
TELEFAX: (510) 655-3542
TELEX: n/a

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3011 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

FEATURE:
NAME/KEY: Duplication
LOCATION: 9
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Lys or Arg"

FEATURE:
NAME/KEY: Duplication
LOCATION: 11
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Asn or Thr"

FEATURE:
NAME/KEY: Duplication
LOCATION: 176
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Ile or Thr"

FEATURE:
NAME/KEY: Duplication
LOCATION: 334
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Met or Val"

FEATURE:
NAME/KEY: Duplication
LOCATION: 2502
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Leu or Phe"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1454
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Cys or Tyr"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1471
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Thr or Ser"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1877
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Glu or Gly"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1948
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Leu or His"

FEATURE:
NAME/KEY: Duplication
LOCATION: 2021
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Gly or Val"

FEATURE:
NAME/KEY: Duplication
LOCATION: 2349
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Thr or Ser"

FEATURE:
NAME/KEY: Duplication
LOCATION: 2385
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Tyr or Phe"

FEATURE:
NAME/KEY: Duplication
LOCATION: 2386
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Ser or Ala"

FEATURE:
NAME/KEY: Duplication
LOCATION: 2502
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Leu or Phe"

FEATURE:
NAME/KEY: Duplication
LOCATION: 848
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Tyr or Asn"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1114
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Pro or Ser"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1117
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Ser or Thr"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1276
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Pro or Leu"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1454
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Cys or Tyr"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1471
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Thr or Ser"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1877
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Glu or Gly"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1948
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Leu or His"

FEATURE:
NAME/KEY: Duplication
LOCATION: 2021
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Gly or Val"

FEATURE:
NAME/KEY: Duplication
LOCATION: 2349
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Thr or Ser"

FEATURE:
NAME/KEY: Duplication
LOCATION: 2385
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Tyr or Phe"

FEATURE:
NAME/KEY: Duplication
LOCATION: 2386
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Ser or Ala"

FEATURE:
NAME/KEY: Duplication
LOCATION: 2502
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Leu or Phe"

```
;
;
; NAME/KEY: Duplication
; LOCATION: 2690
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Arg or Gly"
;
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2921
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Arg or Gly"
;
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2996
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Leu or Pro"
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-232-643-6

Query Match      78.9%; Score 15; DB 16; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.5e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15
Db 571 IGGAGNNTLHCPTDC 585

RESULT 14
US-10-318-200-15
; Sequence 15, Application US/10318200
; Publication No. US20030129746A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: DELPA, ERIK
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: EPITOPES IN VIRAL ENVELOPE PROTEINS AND SPECIFIC
; TITLE OF INVENTION: ANTIBODIES DIRECTED AGAINST THESE EPITOPES: USE FOR
; TITLE OF INVENTION: DETECTION OF HCV VIRAL ANTIGEN IN HOST TISSUE
; FILE REFERENCE: 2551-47
; CURRENT APPLICATION NUMBER: US/10/318,200
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: US/09/645,470
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: PCT/EP99/02154
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: EP 98870060.5
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-318-200-15

Query Match      47.4%; Score 9; DB 16; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTL 9
Db 11 IGGAGNNTL 19

RESULT 15
US-10-128-587A-3
; Sequence 3, Application US/10128587A
; Publication No. US20030152940A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Constructs and methods for expression of recombinant HCV envelope
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: 134 PCT
```

```
;
;
; CURRENT APPLICATION NUMBER: US/10/128,587A
; CURRENT FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 290
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-128-587A-3

Query Match      47.4%; Score 9; DB 12; Length 290;
Best Local Similarity 100.0%; Pred. No. 0.076;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTL 9
Db 188 IGGAGNNTL 196

Search completed: November 21, 2003, 22:19:39
Job time : 18.525 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 154.945 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-85

Perfect score: 19
Sequence: 1 IGGAGNLTLCPTDCRKH 19

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA_Main.*

```

1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US097A_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US097B_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US099A_COMB.pep.*
25: /cgn2_6/ptodata/1/paa/US099B_COMB.pep.*
26: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
27: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
28: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
29: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
30: /cgn2_6/ptodata/1/paa/US104_COMB.pep.*
31: /cgn2_6/ptodata/1/paa/US106_COMB.pep.*
32: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	100.0	19	23	US-09-899-303-85

2	19	100.0	19	23	US-09-899-303A-85	Sequence 85, Appl
3	19	100.0	19	25	US-09-973-025-85	Sequence 85, Appl
4	19	100.0	19	25	US-09-995-791-85	Sequence 85, Appl
5	19	100.0	19	25	US-09-995-808-85	Sequence 85, Appl
6	19	100.0	19	25	US-09-995-860-85	Sequence 85, Appl
7	19	100.0	19	26	US-10-020-510-85	Sequence 85, Appl
8	19	100.0	19	29	US-10-321-798-85	Sequence 85, Appl
9	15	78.9	20	13	US-08-974-685-41	Sequence 41, Appl
10	15	78.9	20	13	US-08-974-690-41	Sequence 41, Appl
11	15	78.9	20	13	US-08-974-690A-41	Sequence 41, Appl
12	15	78.9	20	13	US-08-974-690B-41	Sequence 41, Appl
13	15	78.9	20	13	US-08-974-690C-41	Sequence 41, Appl
14	15	78.9	68	13	US-08-974-690C-250	Sequence 250, App
15	15	78.9	88	8	US-08-471-498-26	Sequence 26, Appl
16	15	78.9	88	8	US-08-471-498-27	Sequence 27, Appl
17	15	78.9	112	8	US-08-436-966-19	Sequence 19, Appl
18	15	78.9	179	8	US-08-403-590B-77	Sequence 77, Appl
19	15	78.9	179	8	US-08-444-112-77	Sequence 77, Appl
20	15	78.9	254	1	PCT-US00-26395-3	Sequence 3, Appl
21	15	78.9	254	18	US-09-407-430-3	Sequence 3, Appl
22	15	78.9	333	30	US-10-445-724-3	Sequence 3, Appl
23	15	78.9	333	32	US-60-409-909-4	Sequence 4, Appl
24	15	78.9	363	1	PCT-US00-26395-2	Sequence 2, Appl
25	15	78.9	363	18	US-09-407-430-2	Sequence 2, Appl
26	15	78.9	414	3	US-07-748-561-2	Sequence 2, Appl
27	15	78.9	414	3	US-07-989-843-32	Sequence 32, Appl
28	15	78.9	414	8	US-08-463-849-32	Sequence 32, Appl
29	15	78.9	414	8	US-08-463-884-32	Sequence 32, Appl
30	15	78.9	414	20	US-09-690-359-32	Sequence 32, Appl
31	15	78.9	454	8	US-08-403-590B-73	Sequence 73, Appl
32	15	78.9	454	8	US-08-444-112-73	Sequence 73, Appl
33	15	78.9	480	8	US-08-471-498-14	Sequence 14, Appl
34	15	78.9	531	21	US-09-721-480-7	Sequence 7, Appl
35	15	78.9	622	3	US-07-748-561-4	Sequence 4, Appl
36	15	78.9	622	3	US-07-989-843-34	Sequence 34, Appl
37	15	78.9	622	8	US-08-463-849-34	Sequence 34, Appl
38	15	78.9	622	8	US-08-463-884-34	Sequence 34, Appl
39	15	78.9	622	20	US-09-690-359-34	Sequence 34, Appl
40	15	78.9	637	1	PCT-US02-20676-4	Sequence 34, Appl
41	15	78.9	637	1	PCT-US02-30423-2	Sequence 2, Appl
42	15	78.9	637	1	PCT-US02-31486-2	Sequence 2, Appl
43	15	78.9	637	27	US-10-187-257-4	Sequence 4, Appl
44	15	78.9	637	28	US-10-265-083-2	Sequence 2, Appl
45	15	78.9	663	6	US-08-282-959-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1

US-09-899-303-85

; Sequence 85, Application US/09899303

; GENERAL INFORMATION:

; APPLICANT: MAERTENS, GEERT

; BOSMAN, FONS

; DE MARTYNOFF, GUY

; BUYSE, MARIE-ANGE

; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

; NUMBER OF SEQUENCES: 111

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: NIXON & VANDERHUYE P.C.

; STREET: 1100 NORTH GLEBE ROAD

; CITY: ARLINGTON

; STATE: VIRGINIA

; COUNTRY: U.S.A.

; ZIP: 22201-4714

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

```
; APPLICATION NUMBER: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 85:
US-09-899-303-85

Query Match          100.0%; Score 19; DB 23; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.4e-12; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0;

QY 1 IGGAGNNTLHCPTDCRKHP 19
    |||||4|||||
Db 1 IGGAGNNTLHCPTDCRKHP 19

RESULT 2
US-09-899-303A-85
; Sequence 85, Application US/09899303A
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303A
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 85:
US-09-899-303-85

Query Match          100.0%; Score 19; DB 23; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.4e-12; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0;

QY 1 IGGAGNNTLHCPTDCRKHP 19
    |||||4|||||
Db 1 IGGAGNNTLHCPTDCRKHP 19

RESULT 2
US-09-899-303A-85
; Sequence 85, Application US/09899303A
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303A
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 85:
US-09-899-303A-85

Query Match          100.0%; Score 19; DB 23; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.4e-12; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0;

QY 1 IGGAGNNTLHCPTDCRKHP 19
    |||||4|||||
Db 1 IGGAGNNTLHCPTDCRKHP 19

RESULT 3
US-09-973-025-85
; Sequence 85, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 85:
US-09-973-025-85

Query Match          100.0%; Score 19; DB 25; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.4e-12; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0;

QY 1 IGGAGNNTLHCPTDCRKHP 19
    |||||4|||||
Db 1 IGGAGNNTLHCPTDCRKHP 19

RESULT 4
US-09-995-791-85
```


; Sequence 85, Application US/09995791
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 85
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-85

Query Match 100.0%; Score 19; DB 25; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDCRKHP 19
| | | | | | | | | | | | | | | | | | | | | |
DB 1 IGGAGNNTLHCPTDCRKHP 19

RESULT 5
US-09-995-808-85
; Sequence 85, Application US/09995808
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 85
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-85

Query Match 100.0%; Score 19; DB 25; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDCRKHP 19
| | | | | | | | | | | | | | | | | | | | | |
DB 1 IGGAGNNTLHCPTDCRKHP 19

RESULT 6
US-09-995-860-85
; Sequence 85, Application US/09995860
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 85
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-85

Query Match 100.0%; Score 19; DB 25; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; Sequence 85, Application US/09995791
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 85
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-85

Query Match 100.0%; Score 19; DB 25; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDCRKHP 19
| | | | | | | | | | | | | | | | | | | | | |
DB 1 IGGAGNNTLHCPTDCRKHP 19

RESULT 5
US-09-995-808-85
; Sequence 85, Application US/09995808
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 85
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-85

Query Match 100.0%; Score 19; DB 25; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDCRKHP 19
| | | | | | | | | | | | | | | | | | | | | |
DB 1 IGGAGNNTLHCPTDCRKHP 19

RESULT 6
US-09-995-860-85
; Sequence 85, Application US/09995860
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 85
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-85

Query Match 100.0%; Score 19; DB 25; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDCRKHP 19
| | | | | | | | | | | | | | | | | | | | | |
DB 1 IGGAGNNTLHCPTDCRKHP 19

RESULT 7
US-10-020-510-85
; Sequence 85, Application US/10020510
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-72
; CURRENT APPLICATION NUMBER: US/10/020,510
; CURRENT FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 85
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-020-510-85

Query Match 100.0%; Score 19; DB 26; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDCRKHP 19
| | | | | | | | | | | | | | | | | | | | | |
DB 1 IGGAGNNTLHCPTDCRKHP 19

RESULT 8
US-10-321-798-85
; Sequence 85, Application US/10321798
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-93
; CURRENT APPLICATION NUMBER: US/10/321,798
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: 60/418,358
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: 10/020,510
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 85
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-321-798-85

Query Match 100.0%; Score 19; DB 29; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDCRKHP 19
| | | | | | | | | | | | | | | | | | | | | |
DB 1 IGGAGNNTLHCPTDCRKHP 19

RESULT 9
US-08-974-685-41
; Sequence 41, Application US/08974685
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; DELEYS, ROBERT
; MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 181
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHUYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,685
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-08-974-685-41

Query Match 78.9%; Score 15; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGAGNNTLHCPTDC 15
| | | | | | | | | | | | | | | | | | | | | |
Db 1 IGGAGNNTLHCPTDC 15

RESULT 10

US-08-974-690-41
; Sequence 41, Application US/08974690
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; APPLICANT: DELEYS, ROBERT
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; TITLE OF INVENTION: HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
; TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 166
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,690
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/635,886

; FILING DATE: 25-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-974-690-41
; Query Match 78.9%; Score 15; DB 13; Length 20;
; Best Local Similarity 100.0%; Pred. No. 2.3e-08;
; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
; Qy 1 IGGAGNNTLHCPTDC 15
; | | | | | | | | | | | | | | | | | | | | | |
; Db 1 IGGAGNNTLHCPTDC 15
; RESULT 11
; US-08-974-690A-41
; ; Sequence 41, Application US/08974690A
; ; GENERAL INFORMATION:
; ; APPLICANT: LEROUX-ROELS, GEERT
; ; APPLICANT: DELEYS, ROBERT
; ; MAERTENS, GEERT
; ; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; ; HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
; ; THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
; ; NUMBER OF SEQUENCES: 177
; ; CORRESPONDENCE ADDRESS:
; ; ADDRESSEE: NIXON & VANDERHUYE P.C.
; ; STREET: 1100 NORTH GLEBE ROAD
; ; CITY: ARLINGTON
; ; STATE: VIRGINIA
; ; COUNTRY: U.S.A.
; ; ZIP: 22201-4714
; ; COMPUTER READABLE FORM:
; ; MEDIUM TYPE: Floppy disk
; ; COMPUTER: IBM PC compatible
; ; OPERATING SYSTEM: PC-DOS/MS-DOS
; ; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; ; CURRENT APPLICATION DATA:
; ; APPLICATION NUMBER: US/08/974,690A
; ; FILING DATE: 19-Nov-1997
; ; CLASSIFICATION: <Unknown>
; ; ATTORNEY/AGENT INFORMATION:
; ; NAME: SADOFF, B. J.
; ; REGISTRATION NUMBER: 36,663
; ; REFERENCE/DOCKET NUMBER: 2752-20
; ; TELECOMMUNICATION INFORMATION:
; ; TELEPHONE: (703) 816-4000
; ; TELEFAX: (703) 816-4100
; ; INFORMATION FOR SEQ ID NO: 41:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 20 amino acids
; ; TYPE: amino acid
; ; STRANDEDNESS: single
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: peptide
; ; SEQUENCE DESCRIPTION: SEQ ID NO: 41:
; ; US-08-974-690A-41
; ; Query Match 78.9%; Score 15; DB 13; Length 20;
; ; Best Local Similarity 100.0%; Pred. No. 2.3e-08;
; ; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

APPLICATION NUMBER: US/08/471,498
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-471-498-26

Query Match 78.9%; Score 15; DB 8; Length 88;

Best Local Similarity 100.0%; Pred. No. 8.2e-08;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 IGGAGNNTLHCPTDC 15

Db 18 IGGAGNNTLHCPTDC 32

Search completed: November 21, 2003, 22:09:53
Job time : 155.945 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.1225 Seconds
(without alignments)
117.536 Million cell updates/sec

Title: US-09-973-025-85

Perfect score: 19

Sequence: 1 IGGAGNNTLHCPTDCRHP 19

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA_New.*

1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	78.9	20	6 US-10-651-165-41	Sequence 41, Appl
2	15	78.9	30	6 US-10-296-734-484	Sequence 484, App
3	15	78.9	68	6 US-10-651-165-250	Sequence 250, App
4	15	78.9	333	1 PCT-US03-19834-3	Sequence 3, Appli
5	15	78.9	333	6 US-10-655-562-4	Sequence 4, Appli
6	15	78.9	637	1 PCT-US03-33610-4	Sequence 4, Appli
7	15	78.9	2010	6 US-10-296-734-814	Sequence 814, App
8	15	78.9	3011	1 PCT-US03-19834-2	Sequence 2, Appli
9	15	78.9	3011	6 US-10-296-734-406	Sequence 406, App
10	15	78.9	5985	6 US-10-296-734-810	Sequence 810, App
11	13	68.4	30	6 US-10-296-734-482	Sequence 482, App
12	11	57.9	68	6 US-10-651-165-252	Sequence 252, App
13	9	47.4	30	6 US-10-685-435-26	Sequence 26, Appl
14	9	47.4	68	6 US-10-651-165-260	Sequence 260, App
15	9	47.4	347	6 US-10-664-391-9	Sequence 9, Appli
16	9	47.4	539	6 US-10-664-391-11	Sequence 11, Appl
17	8	42.1	20	6 US-10-440-390-35	Sequence 35, Appl
18	8	42.1	20	6 US-10-651-165-40	Sequence 40, Appl
19	6	31.6	64	5 US-09-882-583A-14	Sequence 14, Appl
20	6	31.6	595	1 PCT-US03-26354-24	Sequence 24, Appl
21	6	31.6	595	6 US-10-646-308-24	Sequence 24, Appl
22	6	31.6	595	6 US-10-322-696A-156	Sequence 156, App
23	6	31.6	743	1 PCT-US03-24164-47	Sequence 47, Appl
24	6	31.6	1196	5 US-09-897-518A-4602	Sequence 4602, Ap
25	5	26.3	9	6 US-10-428-335-55	Sequence 55, Appl
26	5	26.3	9	6 US-10-428-335-58	Sequence 58, Appl

27	5	26.3	9	6 US-10-428-335-71	Sequence 71, Appl
28	5	26.3	9	6 US-10-428-335-97	Sequence 97, Appl
29	5	26.3	11	6 US-10-651-165-119	Sequence 119, App
30	5	26.3	14	7 US-60-502-656-827	Sequence 827, App
31	5	26.3	22	6 US-10-651-165-130	Sequence 130, App
32	5	26.3	30	7 US-60-507-933-5	Sequence 5, Appli
33	5	26.3	45	6 US-10-653-595-283	Sequence 283, App
34	5	26.3	64	5 US-09-882-583A-9	Sequence 9, Appli
35	5	26.3	64	5 US-09-882-583A-10	Sequence 10, Appl
36	5	26.3	64	5 US-09-882-583A-13	Sequence 13, Appl
37	5	26.3	68	6 US-10-651-165-251	Sequence 251, App
38	5	26.3	68	6 US-10-651-165-253	Sequence 253, App
39	5	26.3	68	6 US-10-651-165-254	Sequence 254, App
40	5	26.3	68	6 US-10-651-165-255	Sequence 255, App
41	5	26.3	68	6 US-10-651-165-256	Sequence 256, App
42	5	26.3	68	6 US-10-651-165-257	Sequence 257, App
43	5	26.3	68	6 US-10-651-165-258	Sequence 258, App
44	5	26.3	68	6 US-10-651-165-259	Sequence 259, App
45	5	26.3	68	6 US-10-651-165-261	Sequence 261, App

ALIGNMENTS

RESULT 1
US-10-651-165-41
; Sequence 41, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-41

Query Match 78.9%; Score 15; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.7e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTLHCPTDC 15
DB 1 IGGAGNNTLHCPTDC 15
|||||

RESULT 2
US-10-296-734-484
; Sequence 484, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2

```
; SEQ ID NO 484
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 39
US-10-296-734-484

Query Match      78.9%; Score 15; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGAGNNTLHCPTDC 15
Db 3 IGGAGNNTLHCPTDC 17

RESULT 3
US-10-651-165-250
; Sequence 250, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; PRIOR FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 250
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-250

Query Match      78.9%; Score 15; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGAGNNTLHCPTDC 15
Db 1 IGGAGNNTLHCPTDC 15

RESULT 4
PCT-US03-19834-3
; Sequence 3, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; TITLE OF INVENTION: METHODS FOR TREATING HYPERLIPIDEMIA
; FILE REFERENCE: IOWA:045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; PRIOR FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Hepatitis C virus
```

PCT-US03-19834-3

```
Query Match      78.9%; Score 15; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 6.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 IGGAGNNTLHCPTDC 15
Db 189 IGGAGNNTLHCPTDC 203
```

RESULT 5

```
US-10-655-562-4
; Sequence 4, Application US/10655562
; GENERAL INFORMATION:
; APPLICANT: BISHOP-HURLEY, SHARON L.
; APPLICANT: SCHMIDT, FRANCIS J.
; APPLICANT: SMITH, ARNOLD L.
; TITLE OF INVENTION: PHAGE-DISPLAY PEPTIDES AS NOVEL ANTIMICROBIAL AGENTS
; FILE REFERENCE: UWMO:022US
; CURRENT APPLICATION NUMBER: US/10/655,562
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: 60,409,909
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-655-562-4
```

```
Query Match      78.9%; Score 15; DB 6; Length 333;
Best Local Similarity 100.0%; Pred. No. 6.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 IGGAGNNTLHCPTDC 15
Db 189 IGGAGNNTLHCPTDC 203
```

RESULT 6

```
PCT-US03-33610-4
; Sequence 4, Application PC/TUS0333610
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: SELBY, Mark
; APPLICANT: PALIARD, Xavier
; TITLE OF INVENTION: ACTIVATION OF HCV-SPECIFIC T-CELLS
; FILE REFERENCE: 2300-1612.60
; CURRENT APPLICATION NUMBER: PCT/US03/33610
; CURRENT FILING DATE: 2003-10-24
; PRIOR APPLICATION NUMBER: 10/281,341
; PRIOR FILING DATE: 2002-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
PCT-US03-33610-4
```

```
Query Match      78.9%; Score 15; DB 1; Length 637;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 IGGAGNNTLHCPTDC 15
Db 399 IGGAGNNTLHCPTDC 413

RESULT 7

US-10-296-734-814

; Sequence 814, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 814
; LENGTH: 2010
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette B
US-10-296-734-814

Query Match 78.9%; Score 15; DB 6; Length 2010;
Best Local Similarity 100.0%; Pred. No. 2.9e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGAGNNTLHCPTDC 15
Db 783 IGGAGNNTLHCPTDC 797

RESULT 8

PCT-US03-19834-2

; Sequence 2, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; FILE REFERENCE: IOWA:045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-2

Query Match 78.9%; Score 15; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 4.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGAGNNTLHCPTDC 15
Db 571 IGGAGNNTLHCPTDC 585

RESULT 9

US-10-296-734-406

; Sequence 406, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A

; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la consensus polyprotein
US-10-296-734-406

Query Match 78.9%; Score 15; DB 6; Length 3011;
Best Local Similarity 100.0%; Pred. No. 4.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGAGNNTLHCPTDC 15
Db 571 IGGAGNNTLHCPTDC 585

RESULT 10

US-10-296-734-810

; Sequence 810, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 810
; LENGTH: 5985
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la savine
US-10-296-734-810

Query Match 78.9%; Score 15; DB 6; Length 5985;
Best Local Similarity 100.0%; Pred. No. 7.5e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGAGNNTLHCPTDC 15
Db 2793 IGGAGNNTLHCPTDC 2807

RESULT 11

US-10-296-794-482 *

; Sequence 482, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 482
; LENGTH: 30

; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 38
US-10-296-734-482

Query Match 68.4%; Score 13; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGAGNNTLHCP 13
Db 18 IGGAGNNTLHCP 30

RESULT 12
US-10-651-165-252
; Sequence 252, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 252
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-252

Query Match 57.9%; Score 11; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 0.00076;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GNNTLHCPDTC 15
Db 5 GNNTLHCPDTC 15

RESULT 13
US-10-685-435-26
; Sequence 26, Application US/10685435
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Multi-mer peptides derived from Hepatitis C Virus
; TITLE OF INVENTION: envelope proteins for diagnostic use and vaccination
; TITLE OF INVENTION: purposes
; FILE REFERENCE: PCT98.75 HCV30
; CURRENT APPLICATION NUMBER: US/10/685,435
; CURRENT FILING DATE: 2003-10-16
; PRIOR APPLICATION NUMBER: US/09/566,266B
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 97870179.5
; PRIOR FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: patenting ver. 2.1
; SEQ ID NO 26
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-685-435-26

Query Match 47.4%; Score 9; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.026;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGAGNNTL 9
Db 11 IGGAGNNTL 19

RESULT 14
US-10-651-165-260
; Sequence 260, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 260
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-260

Query Match 47.4%; Score 9; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 0.052;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IGGAGNNTL 9
Db 1 IGGAGNNTL 9

RESULT 15
US-10-664-391-9
; Sequence 9, Application US/10664391
; GENERAL INFORMATION:
; APPLICANT: Donnelly, John J.
; APPLICANT: Liu, Margaret A.
; APPLICANT: Shiver, John W.
; APPLICANT: Fu, Tong-Ming
; TITLE OF INVENTION: SYNTHETIC HEPATITIS C GENES
; FILE REFERENCE: 19732VPCA
; CURRENT APPLICATION NUMBER: US/10/664,391
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: PCT/US97/09884
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/033,534
; PRIOR FILING DATE: 1996-12-20
; PRIOR APPLICATION NUMBER: 60/020,494
; PRIOR FILING DATE: 1996-06-11
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-10-664-391-9

Query Match 47.4%; Score 9; DB 6; Length 347;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ICGAGNNTL 9
| | | | |
Db 189 ICGAGNNTL 197

Search completed: November 21, 2003, 22:12:58
Job time : 9.1225 secs

0 1

0 1

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:46:27 ; Search time 8.7875 Seconds
(without alignments)
207.932 Million cell updates/sec

Title: US-09-973-025-85
Perfect score: 19
Sequence: 1 IGGAGNNTLHCPTDCRHP 19
Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues
Word size : 0
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	78.9	3011	1 GNVWC3	genome polyprotein
2	11	57.9	640	2 JQ1584	genome polyprotein
3	11	57.9	3011	1 S40770	genome polyprotein
4	9	47.4	716	2 JQ1366	polyprotein - hepa
5	9	47.4	782	2 S19876	genome polyprotein
6	9	47.4	782	2 S19875	genome polyprotein
7	9	47.4	3010	1 S18030	genome polyprotein
8	8	42.1	3262	2 AH2137	hypothetical prote
9	7	36.8	257	2 S74478	hypothetical prote
10	7	36.8	364	1 B64346	cell division prot
11	7	36.8	369	2 G64224	cell division prot
12	7	36.8	375	1 T44848	cell division prot
13	7	36.8	375	2 H84179	cell division prot
14	7	36.8	380	2 S73845	cell division prot
15	7	36.8	381	2 A69091	cell division prot
16	7	36.8	398	2 H70393	hemolysin - Aquife
17	7	36.8	403	2 F64377	cell division prot
18	7	36.8	1023	1 LERCA	hemolysin A - Esch
19	6	31.6	174	2 D70979	probable purE prot
20	6	31.6	191	2 A43597	pilin type F9 prec
21	6	31.6	305	2 T28005	hypothetical prote
22	6	31.6	313	2 T23289	hypothetical prote
23	6	31.6	313	2 H96037	alpha-galactoside
24	6	31.6	351	2 H72328	cell division prot
25	6	31.6	371	2 E75494	cell division prot
26	6	31.6	377	2 AD1937	permease protein o
27	6	31.6	427	2 F64064	to1B protein - Hae
28	6	31.6	444	2 S40980	hypothetical prote
29	6	31.6	446	1 UB2PG	tubulin gamma chai

30	6	31.6	449	2 H69423	branched-chain ami
31	6	31.6	451	1 UBHUG	tubulin gamma chai
32	6	31.6	451	1 UBXLG	tubulin gamma chai
33	6	31.6	457	2 T08419	tubulin gamma chai
34	6	31.6	462	2 S53084	gamma-tubulin - Eu
35	6	31.6	468	2 T08057	tubulin gamma chai
36	6	31.6	468	2 T07904	tubulin gamma chai
37	6	31.6	469	2 S31727	tubulin gamma chai
38	6	31.6	469	2 S44193	tubulin gamma chai
39	6	31.6	472	2 S39553	tubulin gamma chai
40	6	31.6	474	2 T50558	tubulin gamma-2 ch
41	6	31.6	474	2 T47957	tubulin gamma-1 ch
42	6	31.6	475	1 UBFFG	tubulin gamma chai
43	6	31.6	493	2 AD2354	hypothetical prote
44	0	31.6	535	2 AD2107	serine/threonine k
45	6	31.6	556	2 AH1981	hypothetical prote

ALIGNMENTS

RESULT 1

GNWVC3
genome polyprotein - hepatitis C virus (strain HCV-1)
N:Contains: capsid protein C; envelope protein M; hepatitisvirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 30-Sep-1992 Sequence revision 30-Sep-1992 #text_change 19-Jan-2001
C:Accession: A39166; PQ0403; PQ0404
R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A:Title: Genetic organization and diversity of the hepatitis C virus.
A:Reference number: A39166; MUID:91172826; PMID:1848704
A:Accession: A39166
A:Molecule type: mRNA
A:Residues: 1-3011 <CHO>
A:Cross-references: GB:M62321; NID:G329873; PIDN:AAA45676.1; PID:G329874
J:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.I
J. Gen. Virol. 73, 1131-1141, 1992
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A:Reference number: PQ0393; MUID:92268871; PMID:1316939
A:Accession: PQ0403
A:Molecule type: Genomic RNA
A:Residues: 1577-1633 <CHA>
A:Cross-references: DDBJ:D10128
A:Experimental source: Isolates E-b16
A:Accession: PQ0404
A:Status: preliminary
A:Molecule type: Genomic RNA
A:Residues: 1577-1633 <CH2>
A:Experimental source: isolates E-b17
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F:115-Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein NS1 #status predicted <MEE>
F:390-729/Product: nonstructural protein NS2 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitisvirin #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (p-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,556,576,623,645,1213,1255,2041,2077,22

Query Match 78.9% Score 15; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 8.7e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IGGAGNNTLHCPTDC 15
|||||

Db 571 IGGAGNNTLHCPTDC 585

RESULT 2

JQ1584

genome polyprotein - hepatitis C virus (strain U.K.) (fragment)

N:Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural protein

C:Species: hepatitis C virus

C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000

C:Accession: JQ1584

R:Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.

J. Gen. Virol. 73, 1521-1525, 1992

A:Title: Cloning and sequencing of the structural region and expression of putative core

A:Reference number: JQ1584; MUID:92300349; PMID:1318944

A:Accession: JQ1584

A:Molecule type: genomic RNA

A:Residues: 1-640 <KUM>

A:Cross-references: GB:X84079; NID:g643119; PIDN:CAA58888.1; PID:g643120

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; polypro

F:1-191/Product: core protein C #status predicted <CPC>

F:192-389/Product: envelope protein E1 #status predicted <EE1>

F:390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predicted <

F:196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn) (cova

Query Match 57.9%; Score 11; DB 2; Length 640;

Best Local Similarity 100.0%; Pred. No. 0.0004;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GNNTLHCPTDC 15

Db 575 GNNTLHCPTDC 585

RESULT 3

S40770

genome polyprotein - hepatitis C virus

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain U.K.) (fragment)

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001

C:Accession: S40770; PCL285

R:Okamoto, H.

submitted to the EMBL Data Library, March 1992

A:Reference number: S40770

A:Accession: S40770

A:Molecule type: genomic RNA

A:Residues: 1-3011 <OKA>

A:Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587

R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,

Jpn. J. Exp. Med. 60, 167-177, 1990

A:Title: The 5'-terminal sequence of the hepatitis C virus genome.

A:Reference number: PCL284; MUID:91013116; PMID:2170712

A:Accession: PCL285

A:Molecule type: genomic RNA

A:Residues: 1-513 <OK2>

A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512

A:Experimental source: isolate HC-J1

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin

F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEP>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: nonstructural protein NS3

F:1230-1237/Region: hepatitis C virus genome polyprotein

F:1312-1317/Region: nucleotide-binding motif A (P-loop)

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 57.9%; Score 11; DB 1; Length 3011;

Best Local Similarity 100.0%; Pred. No. 0.0014;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GNNTLHCPTDC 15

Db 575 GNNTLHCPTDC 585

RESULT 4

JQ1366

polyprotein - hepatitis C virus (French isolate) (fragments)

C:Species: hepatitis C virus

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000

C:Accession: JQ1366

R:Kremsdorf, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.

J. Gen. Virol. 72, 2557-2561, 1991

A:Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication

A:Reference number: JQ1366; MUID:92013977; PMID:1655961

A:Accession: JQ1366

A:Molecule type: genomic RNA

A:Residues: 1-716 <KRE>

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: glycoprotein; polyprotein

F:84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #strat

Query Match 47.4%; Score 9; DB 2; Length 716;

Best Local Similarity 100.0%; Pred. No. 0.055;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTL 9

Db 238 IGGAGNNTL 246

RESULT 5

S19876

genome polyprotein - hepatitis C virus (isolate JKS) (fragment)

N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein

C:Species: hepatitis C virus

A:Variety: isolate JKS

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000

C:Accession: S19876

R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.

submitted to the EMBL Data Library, September 1991

A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso

A:Reference number: S18029

A:Accession: S19876

A:Molecule type: genomic RNA

A:Residues: 1-782 <HON>

A:Cross-references: EMBL:X61595; NID:g59486; PIDN:CAA43792.1; PID:g59487

A:Experimental source: isolate JKS

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural

F:1-191/Product: core protein #status predicted <MAT1>

F:192-383/Product: envelope protein 1 #status predicted <MAT2>

F:384-733/Product: NS1/E2 protein #status predicted <MAR3>

F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 47.4%; Score 9; DB 2; Length 782;

Best Local Similarity 100.0%; Pred. No. 0.059;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNNTL 9

Db 571 IGGAGNNTL 579

RESULT 6

S19875

genome polyprotein - hepatitis C virus (isolate JK3) (fragment)

N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein

C:Species: hepatitis C virus

A:Variety: isolate JK3
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
 C:Accession: S19875
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus isolate JK3
 A:Reference number: S18029
 A:Accession: S19875
 A:Molecule type: genomic RNA
 A:Residues: 1-782 <HON>
 A:Cross-references: EMBL:X61592; NID:G59482; PIDN:CRAA43789.1; PID:G59483
 A:Experimental source: isolate JK3
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
 F:1-191/Product: core protein #status predicted <MAT1>
 F:192-383/Product: envelope protein 1 #status predicted <MAT2>
 F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
 F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 47.4%; Score 9; DB 2; Length 782;
 Best Local Similarity 100.0%; Pred. No. 0.059;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ICGAGNNTL 9
 |||||
 DB 571 ICGAGNNTL 579

RESULT 7
 S18030
 genome polyprotein - hepatitis C virus (isolate JK1)
 N:Contains: capsid protein C; envelope protein M; Hepacivirus protein NS5 (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)
 C:Species: hepatitis C virus
 A:Variety: isolate JK1
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
 C:Accession: S18030; S33570; A48332; S18029
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient
 A:Reference number: S18028
 A:Accession: S18030
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <HON>
 A:Cross-references: EMBL:X61596; NID:G59478; PIDN:CRAA43793.1; PID:G59479
 A:Experimental source: isolate JK1 from an individual
 R:Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.
 Arch. Virol. 128, 163-169, 1993
 A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolated
 A:Reference number: A48332; MUID:93119270; PMID:8380322
 A:Accession: S33570
 A:Molecule type: genomic RNA
 A:Residues: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HON>
 A:Cross-references: EMBL:X61591
 A:Note: this sequence is inconsistent with the nucleotide translation
 A:Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320 as Trp, and TTC for residue 771 as Ser
 A:Note: sequence extracted from NCBI backbone (NCBI:121747, NCBI:121748)
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; Glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEP>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: Hepacivirus #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:1196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As

Query Match 47.4%; Score 9; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.17;
 Matches 0 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ICGAGNNTL 9
 |||||
 DB 571 ICGAGNNTL 579

RESULT 8
 AH2137
 hypothetical protein all2655 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AH2137
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AH2137
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-3262 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA074354.1; PID:GI7131748; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all2655

Query Match 42.1%; Score 8; DB 2; Length 3262;
 Best Local Similarity 100.0%; Pred. No. 2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGAGNNTL 9
 |||||
 DB 3113 GGAGNNTL 3120

RESULT 9
 S74478
 hypothetical protein slr1115 - Synechocystis sp. (strain PCC 6803)
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
 C:Accession: S74478
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S74478
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-257 <KAN>
 A:Cross-references: EMBL:D90899; GB:AB001339; NID:G1651650; PIDN:BA016630.1; PID:dl01736
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Genetics:
 A:Start codon: GTG

Query Match 36.8%; Score 7; DB 2; Length 257;
 Best Local Similarity 100.0%; Pred. No. 3;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GAGNNTL 9
 |||||
 DB 80 GAGNNTL 86

RESULT 10
 B64346

cell division protein FtsZ homolog MJ0370 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 04-Oct-1996 #text_change 21-Jul-2000
C:Accession: B64346
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Aron, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: B64346
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-364 <BUL>
A:CROSS-references: GB:U67490; GB:L77117; NID:g2826276; PIDN:AAB98359.1; PID:g1591077; T
C:Comment: This protein is the nearer of two homologs in the archaeon Methanococcus jan
f self-assembly and required for initiating the partition between compartments during ce
C:Genetics:
A:Map position: FOR336314-337408
A:Start codon: GTG
C:Superfamily: cell division protein ftsz
C:Keywords: cell division; GTP binding
F:132-138/Region: tubulin/FtsZ GTP/GDP-binding (G-G-G-T-G-[ST]-G) motif

Query Match 36.8%; Score 7; DB 1; Length 364;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGAGNNT 8
DB 46 GGAGNNT 52
|||||

RESULT 11
G64224
cell division protein FtsZ - Mycoplasma genitalium
C:Species: Mycoplasma genitalium
C:Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 07-Dec-1999
C:Accession: G64224
R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.
C.A.; Venter, J.C.
Science 270, 397-403, 1995
A:Title: The minimal gene complement of Mycoplasma genitalium.
A:Reference number: A64200; MUID:96026346; PMID:7569993
A:Accession: G64224
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-369 <TIGR>
A:CROSS-references: GB:U39700; GB:L43967; NID:g1045906; PID:g1045912; TIGR:MG224
A:Experimental source: strain G-37
C:Genetics:
A:Genetic code: SGC3
C:Superfamily: cell division protein ftsz
C:Keywords: cell division; GTP binding
F:117-123/Region: tubulin/FtsZ GTP/GDP-binding (G-G-G-T-G-[ST]-G) motif #status atypical

Query Match 36.8%; Score 7; DB 2; Length 369;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNN 7
DB 25 IGGAGNN 31
|||||

RESULT 12
T44848
cell division protein ftsz [validated] - Halobacterium salinarum
N:Alternate names: GTP-binding protein ftsz
C:Species: Halobacterium salinarum
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000

C:Accession: T44848
R:Margolin, W.; Wang, R.; Kumar, M.
J. Bacteriol. 178, 1320-1327, 1996
A:Title: Isolation of an ftsZ homolog from the archaeobacterium Halobacterium salinarum:
A:Reference number: 222858; MUID:96200101; PMID:8631708
A:Accession: T44848
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-375 <MAR>
C:Genetics:
A:Gene: ftsZ
C:Function:
A:Description: involved in cell division; overexpression of ftsZ induces significant mor
C:Superfamily: cell division protein ftsz
C:Keywords: cell division; GTP binding

Query Match 36.8%; Score 7; DB 1; Length 375;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGAGNNT 8
DB 21 GGAGNNT 27
|||||

RESULT 13
H84179
cell division protein [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: H84179
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hsu, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: H84179
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-375 <STO>
A:CROSS-references: GB:AE004437; NID:g10579838; PIDN:AAG18804.1; GSPDB:GN00138
C:Genetics:
A:Gene: ftsZ
C:Superfamily: cell division protein ftsz

Query Match 36.8%; Score 7; DB 2; Length 375;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGAGNNT 8
DB 21 GGAGNNT 27
|||||

RESULT 14
S73845
cell division protein ftsZ - Mycoplasma pneumoniae (strain ATCC 29342)
N:Alternate names: hypothetical protein P10_orf380
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
C:Accession: S73845
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73327; MUID:97105885; PMID:8948633
A:Accession: S73845
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-380 <HIM>

A;Cross-references: EMBL:AB000051; GB:U00089; NID:g1674211; PIDN:AAB96167.1; PID:g1674211
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C;Genetics:
A;Gene: ftsZ
A;Genetic code: SGC3
C;Superfamily: cell division protein ftsZ
F;117-123/Region: tubulin/FtsZ GTP/GDP-binding (G-G-G-T-G-[ST]-G) motif #status atypical

Query Match 36.8%; Score 7; DB 2; Length 380;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNN 7
|||
DB 25 IGGAGNN 31

RESULT 15

AE9091
cell division protein FtsZ - Methanobacterium thermoautotrophicum (strain Delta H)
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-Aug-1999
C;Accession: AE9091
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
Kl, S.; Church, G.M.; Daniele, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: func
A;Reference number: AE9000; MUID:98037514; PMID:9371463
A;Accession: AE9091
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-381 <MTH>
A;Cross-references: GB:AE000925; GB:AE000666; NID:g2622791; PIDN:AAB86148.1; PID:g262280
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH1676
A;Start codon: TTG
C;Superfamily: cell division protein ftsZ
F;133-139/Region: tubulin/FtsZ GTP/GDP-binding (G-G-G-T-G-[ST]-G) motif

Query Match 36.8%; Score 7; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGAGNNT 8
|||
DB 47 GGAGNNT 53

Search completed: November 21, 2003, 21:11:35
Job time : 8.7875 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6

Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:52 ; Search time 4.6075 Seconds
(without alignments)
193.925 Million cell updates/sec

Title: US-09-973-025-85

Perfect score: 19

Sequence: 1 IGGAGNNTLHCPTDCRKH 19

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	78.9	3011	1 POLG_HCV1	P26664 h genome po
2	7	36.8	364	1 FT21_METJA	Q57816 methanococ
3	7	36.8	369	1 FTSZ_MYCGE	P47466 mycoplasma
4	7	36.8	375	1 FTSZ_HALN1	Q48290 halobacteri
5	7	36.8	380	1 FTSZ_MYCPN	P75464 mycoplasma
6	7	36.8	380	1 FT22_METJA	Q58039 methanococ
7	7	36.8	381	1 FTSZ_METTH	Q27712 methanobact
8	7	36.8	1023	1 HLVI_ECOLI	P09983 escherichia
9	6	31.6	174	1 PUR6_MYCTU	P96880 mycobacteri
10	6	31.6	351	1 FTSZ_THEMEA	O08398 thermotoga
11	6	31.6	421	1 TBG3_WAIZE	Q41874 zea mays (m
12	6	31.6	427	1 TOLB_HABIN	P44677 haemophilus
13	6	31.6	444	1 TBG_CABEL	P34475 caenorhabdi
14	6	31.6	446	1 TBG_SCHJP	Q9Y982 schizosacch
15	6	31.6	446	1 TBG_SCHPO	P25295 schizosacch
16	6	31.6	451	1 TBG1_HUMAN	P23258 homo sapien
17	6	31.6	451	1 TBG1_MOUSE	Q92310 mus musculu
18	6	31.6	451	1 TBG2_HUMAN	Q9nrh3 homo sapien
19	6	31.6	451	1 TBG2_MOUSE	Q8vck3 mus musculu
20	6	31.6	451	1 TBG_XENLA	P23330 xenopus lae
21	6	31.6	452	1 TBG_PLAFO	P34787 plasmodium
22	6	31.6	457	1 TBG2_DROME	P42271 drosophila
23	6	31.6	461	1 TBG1_EUPOC	P34786 eulotes oc
24	6	31.6	461	1 TBG2_EUPOC	P90548 eulotes oc
25	6	31.6	461	1 TBG_EUPAE	P54402 eulotes ae
26	6	31.6	462	1 TBSP_DROME	Q9vrx3 drosophila
27	6	31.6	462	1 TBG1_EUPCR	P54403 eulotes cr
28	6	31.6	468	1 TBG_CHLRE	Q39582 chlamydomon
29	6	31.6	469	1 TBG1_WAIZE	Q41807 zea mays (m
30	6	31.6	469	1 TBG2_WAIZE	Q41808 zea mays (m
31	6	31.6	469	1 TBG2_ORYSA	Q49068 oryza sativ
32	6	31.6	469	1 TBG_USTVI	P32348 ustilago vi
33	6	31.6	472	1 TBG_ANEPH	P34785 anemia phyl

RESULT 1

ID	POLG_HCV1	STANDARD;	PRT;	3011 AA.
AC	P26664;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].			
OS	Hepatitis C virus (isolate 1) (HCV).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.			
OX	NCBI_TaxID=11104;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91172826; PubMed=1848704;			
RA	Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;			
RT	"Genetic organization and diversity of the hepatitis C virus.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).			
CC	!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.			
CC	!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA)(N).			
CC	!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.			
CC	!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; M62321; AAA45676.1; -			
DR	PIR; A39166; GNMVCS3			
DR	PDB; 1A1V; 16-FEB-99.			
DR	PDB; 1HE1; 25-NOV-98.			
DR	MEROPS; S29.001; -			
DR	MEROPS; U39.001; -			
DR	InterPro; IPR001410; DEAD.			
DR	InterPro; IPR002522; HCV_capsid.			

ALIGNMENTS

DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR007095; RNA_pol_DS_Ps.
 DR InterPro; IPR007094; RNA_pol_PsVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00598; Viral_RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC2; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3011
 FT CHAIN 3012 347
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 476 476
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2364 2364
 FT CARBOHYD 2789 2789
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;
 Query Match 78.9%; Score 15; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 1.2e-08;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 IGGAGNTLHCPTDC 15

Db 571 IGGAGNTLHCPTDC 585
 RESULT 2
 FTZ1_METJA
 ID FTZ1_METJA STANDARD; PRT; 364 AA.
 AC Q57816;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cell division protein ftsZ homolog 1.
 GN FTSZ1 OR MJ0370.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcales; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A. DSM 2661 / ATCC 43067;
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Sult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RL Science 273:1058-1073 (1996).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RX MEDLINE=98089046; PubMed=9428770;
 RA Loewe J., Amos L.A.;
 RT "Crystal structure of the bacterial cell-division protein ftsZ.";
 RL Nature 391:203-206 (1998).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 23-356.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=99160450; PubMed=10049809;
 RA Loewe J.;
 RT "Crystal structure determination of FtsZ from Methanococcus
 jannaschii.";
 RL J. Struct. Biol. 124:235-243 (1998).
 CC -!- FUNCTION: This protein is essential to the cell-division process.
 CC It seems to assemble into a dynamic ring on the inner surface of
 CC the cytoplasmic membrane at the place where division will occur,
 CC and the formation of the ring is the signal for septation to
 CC begin. Binds to and hydrolyzes GTP (By similarity).
 CC -!- SUBUNIT: Aggregates to form a ring-like structure (By similarity).
 CC -!- DOMAIN: CONSISTS OF TWO GLOBULAR DOMAINS WITH A CONNECTING CORE
 CC HELIX.
 CC -!- SIMILARITY: BELONGS TO THE FTSZ FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL; U67490; AAB98359.1; -;
 DR PIR; B64346; B64346.
 DR PDB; 1FSZ; 24-JUN-98.
 DR TIGR; MJ0370; -;
 DR InterPro; IPR000158; FtsZ.
 DR InterPro; IPR003008; Tubulin_FtsZ.
 DR Pfam; PF00091; tubulin; 1.
 DR Pfam; PF03953; tubulin.C; 1.
 DR PRINTS; PR00423; CELLDIVISITSZ.

DR TIGRFAMs; TIGR00065; ftsZ, 1.
 DR PROSITE; PS01134; ftsZ_1; 1.
 DR PROSITE; PS01135; ftsZ_2; 1.
 KW Cell division; Septation; GTP-binding; Multigene family; 3D-structure;
 KW Complete proteome.
 FT NP_BIND 130 138 GTP.
 FT HELIX 24 34
 FT STRAND 40 45
 FT STRAND 46 59
 FT TURN 62 63
 FT TURN 65 70
 FT STRAND 72 72
 FT STRAND 73 77
 FT TURN 78 78
 FT STRAND 83 86
 FT STRAND 89 91
 FT STRAND 94 95
 FT TURN 98 99
 FT TURN 101 110
 FT TURN 111 111
 FT HELIX 112 119
 FT TURN 120 121
 FT STRAND 124 130
 FT TURN 131 132
 FT HELIX 135 149
 FT TURN 150 151
 FT STRAND 153 160
 FT HELIX 163 165
 FT HELIX 167 182
 FT TURN 183 183
 FT STRAND 186 191
 FT HELIX 192 197
 FT TURN 198 198
 FT HELIX 204 227
 FT HELIX 236 243
 FT TURN 244 245
 FT STRAND 248 256
 FT HELIX 261 271
 FT TURN 273 274
 FT TURN 279 280
 FT STRAND 285 281
 FT TURN 293 294
 FT HELIX 297 308
 FT TURN 309 310
 FT TURN 313 314
 FT TURN 317 323
 FT STRAND 328 329
 FT STRAND 331 337
 FT TURN 341 343
 FT STRAND 344 347
 FT TURN 348 349
 FT STRAND 350 353
 SQ SEQUENCE 364 AA; 38924 MW; 3B9386A5D2FCA107 CRC64;
 Query Match 36.8%; Score 7; DB 1; Length 364;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GGAGNNT 8
 DB 46 GGAGNNT 52
 RESULT 3
 FTSZ_MYCGE STANDARD; PRT; 369 AA.
 AC P47466;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cell division protein ftsZ.
 OS Mycoplasma genitalium.
 NCBI_TaxID=64091, 2242;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRC-1;

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE=96026346; PubMed=7569993;
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
 RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.-P., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
 RT "The minimal gene complement of Mycoplasma genitalium.";
 RL Science 270:397-403(1995).
 CC -!- FUNCTION: This protein is essential to the cell-division process.
 CC It seems to assemble into a dynamic ring on the inner surface of
 CC the cytoplasmic membrane at the place where division will occur,
 CC and the formation of the ring is the signal for septation to
 CC begin. Binds to and hydrolyzes GTP (by similarity).
 CC -!- SUBUNIT: Aggregates to form a ring-like structure (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Assembles at the inner surface
 CC of the cytoplasmic membrane (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE FTSZ FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U39702; AAC71445.1; -.
 DR PIR; G64224; G64224.
 DR HSP; Q57816; ftsZ.
 DR TIGR; MG224; -.
 DR InterPro; IPR000158; ftsZ.
 DR InterPro; IPR003008; Tubulin_FtsZ.
 DR Pfam; PF00091; tubulin; 1.
 DR PRINTS; PR00423; CELLDVIFTSZ.
 DR TIGRFAMs; TIGR00065; ftsZ; 1.
 DR PROSITE; PS01134; ftsZ_1; 1.
 DR PROSITE; PS01135; ftsZ_2; 1.
 KW Cell division; Septation; GTP-binding; Complete proteome.
 FT NP_BIND 115 123 GTP (POTENTIAL).
 SQ SEQUENCE 369 AA; 41603 MW; 6173D93AA0FCF5A0 CRC64;
 Query Match 36.8%; Score 7; DB 1; Length 369;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 IGGAGNN 7
 DB 25 IGGAGNN 31
 RESULT 4
 FTSZ_HALN1 STANDARD; PRT; 375 AA.
 ID FTSZ_HALN1
 AC Q48290; Q9H8K0;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cell division protein ftsZ homolog.
 GN FTSZ OR ftsZ2 OR VNG0192G.
 OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081), and
 OS Halobacterium salinarum.
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 OX NCBI_TaxID=64091, 2242;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRC-1;

```

RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=ATCC 29342 / M129;
RX      MEDLINE=97105885; PubMed=8948633;
RA      Himmelreich R., Hilbert H., Plagens H., Pirkle E., Li B.-C.,
RA      Herrmann R.;
RT      "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT      pneumoniae.";
RL      Nucleic Acids Res. 24:4420-4449(1996).
CC      -!- FUNCTION: This protein is essential to the cell-division process.
CC      It seems to assemble into a dynamic ring on the inner surface of
CC      the cytoplasmic membrane at the place where division will occur,
CC      and the formation of the ring is the signal for septation to
CC      begin. Binds to and hydrolyzes GTP (By similarity).
CC      -!- SUBUNIT: Aggregates to form a ring-like structure (By similarity).
CC      -!- SUBCELLULAR LOCATION: Cytoplasmic. Assembles at the inner surface
CC      of the cytoplasmic membrane (By similarity).
CC      -!- SIMILARITY: BELONGS TO THE FTSZ FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AE000051; BAB96167.1; .
DR      PIR; S73845; S73845. FtsZ.
DR      InterPro; IPR000158; FtsZ.
DR      InterPro; IPR003008; Tubulin_FtsZ.
DR      Pfam; PF00091; tubulin; 1.
DR      PRINTS; PR00423; CELLDVIFTSZ.
DR      TIGRFS; TIGR00065; ftsz; 1.
DR      PROSITE; PS01134; FTSZ_1; 1.
DR      PROSITE; PS01135; FTSZ_2; 1.
DR      Cell division; Septation; GTP-binding; Complete proteome.
KW      NP BIND 115 123 GTP (POTENTIAL).
FT      SEQUENCE 380 AA; 42797 MW; 6C3B531C63F4997E CRC64;
SQ
Query Match 36.8%; Score 7; DB 1; Length 380;
Best Local Similarity 100.0%; Pred.No.1.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IGGAGNN 7
DB 25 IGGAGNN 31
|||||
RESULT 6
FT22_METJ2
ID FT22_METJ2 STANDARD; PRT; 380 AA.
AC Q58039;
DT 11-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cell division protein ftsZ homolog 2.
GN FTSZ2 OR MJ0622.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2561 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Overbeek A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Kierulff A., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.D., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,

```

RA Klenk H.-P., Frazer C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, *Methanococcus*
 RL *jannaschii*.";
 CC Science 273:1058-1073(1996).
 CC -!- FUNCTION: This protein is essential to the cell-division process.
 CC It seems to assemble into a dynamic ring on the inner surface of
 CC the cytoplasmic membrane at the place where division will occur,
 CC and the formation of the ring is the signal for septation to
 CC begin. Binds to and hydrolyzes GTP (By similarity).
 CC -!- SUBUNIT: Aggregates to form a ring-like structure (By similarity).
 CC -!- DOMAIN: CONSISTS OF TWO GLOBULAR DOMAINS WITH A CONNECTING CORE
 CC -!- SIMILARITY: BELONGS TO THE FTSZ FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U67510; AAB98617.1; -.
 DR HSP; Q57816; 1FSZ.
 DR TIGR; MJ0622; -.
 DR InterPro; IPR000158; FtsZ.
 DR InterPro; IPR003008; Tubulin_FtsZ.
 DR Pfam; PF00091; tubulin; 1.
 DR Pfam; PF03953; tubulin_C; 1.
 DR PRINTS; PR00423; CELLDVISFTSZ.
 DR TIGRFAMs; TIGR00065; ftsz; 1.
 DR PROSITE; PS01134; FTSZ_1; 1.
 DR PROSITE; PS01135; FTSZ_2; 1.
 DR Cell division; Septation; GTP-binding; Multigene family;
 KW Complete proteome. 138 GTP (POTENTIAL).
 FT NP_BIND 130 138
 SQ SEQUENCE 380 AA; 40658 MW; 2AJAD01CCB54FB8C CRC64;
 Query Match 36.8%; Score 7; DB 1; Length 380;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GGAGNNT 8
 Db 46 GGAGNNT 52
 |||||
 RESULT 7
 FTSZ_METTH
 ID FTSZ_METTH STANDARD; PRT; 381 AA.
 AC O27712;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cell division protein ftsz.
 GN FTSZ OR MTH1676.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 OX NCBI_TaxID=187420;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Delta H;
 RX MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwni N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RT "Complete genome sequence of *Methanobacterium thermoautotrophicum*
 RT *deltaH*: functional analysis and comparative genomics.";

RL J. Bacteriol. 179:7135-7155(1997).
 CC -!- FUNCTION: This protein is essential to the cell-division process.
 CC It seems to assemble into a dynamic ring on the inner surface of
 CC the cytoplasmic membrane at the place where division will occur,
 CC and the formation of the ring is the signal for septation to
 CC begin. Binds to and hydrolyzes GTP (By similarity).
 CC -!- SUBUNIT: Aggregates to form a ring-like structure (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE FTSZ FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE000925; AAB86148.1; -.
 DR PIR; A69091; A69091.
 DR HSP; Q57816; 1FSZ.
 DR InterPro; IPR000158; FtsZ.
 DR InterPro; IPR003008; Tubulin_FtsZ.
 DR Pfam; PF00091; tubulin; 1.
 DR Pfam; PF03953; tubulin_C; 1.
 DR PRINTS; PR00423; CELLDVISFTSZ.
 DR TIGRFAMs; TIGR00065; ftsz; 1.
 DR PROSITE; PS01134; FTSZ_1; 1.
 DR PROSITE; PS01135; FTSZ_2; 1.
 DR Cell division; Septation; GTP-binding; Complete proteome.
 KW NP_BIND 131 139
 FT NP_BIND 131 139
 SQ SEQUENCE 381 AA; 40619 MW; 8E96863DAF012EE CRC64;
 Query Match 36.8%; Score 7; DB 1; Length 381;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GGAGNNT 8
 Db 47 GGAGNNT 53
 |||||
 RESULT 8
 HLY1_ECOLI
 ID HLY1_ECOLI STANDARD; PRT; 1023 AA.
 AC P09983;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hemolysin, chromosomal.
 GN HLYA.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=j96 / Serotype O4;
 RX MEDLINE=85234404; PubMed=3891743;
 RA Felmei T., Pellett S., Welch R.A.;
 RT "Nucleotide sequence of an *Escherichia coli* chromosomal hemolysin.";
 RL J. Bacteriol. 163:94-105(1985).
 RN [2]
 RP SEQUENCE OF 1-44 FROM N.A.
 RC STRAIN=2001;
 RX MEDLINE=85258115; PubMed=3894051;
 RA Nicaud J.-M., Mackman N., Gray L., Holland I.B.;
 RT "Characterisation of HlyC and mechanism of activation and secretion
 RT of haemolysin from *E. coli* 2001.";
 RL FEBS Lett. 187:339-344(1985).
 CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
 CC cell membranes and cause cell rupture by mechanisms not clearly
 CC defined.
 CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- DOMAIN: The Gly-rich region is probably involved in binding
 CC calcium, which is required for target cell-binding or cytolytic
 CC activity.
 CC -!- DOMAIN: The three transmembrane domains are believed to be
 CC involved in pore formation by the cytotoxin.
 CC -!- PTM: Palmitoylated by hlyC. The toxin only becomes active when
 CC modified.
 CC -!- MISCELLANEOUS: The hemolysin of E.coli is produced predominantly
 CC by strains causing extraintestinal infections, such as those of
 CC the urinary tract.
 CC -!- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M10133; AAA23975.1; -;
 CC EMBL; X02768; CAA26546.1; -;
 CC PIR; A24433; LEECA.
 CC InterPro; IPR001343; Hemlysn_Ca_bind.
 CC InterPro; IPR003995; RTXa.
 CC Pfam; PF00353; hemolysinCbind; 6.
 CC Pfam; PF02382; RTX; 1.
 CC PRINTS; PR00313; CABNDNGRPT.
 CC PRINTS; PR01488; RTXTOXINA.
 CC PROSITE; PS00330; HEMOLYSIN CALCIUM; 4.
 CC Hemolysin; Toxin; Cytolysis; Repeat; Calcium; Transmembrane;
 CC Lipoprotein; Palmitate.

FT TRANSMEM 237 259 POTENTIAL.
 FT TRANSMEM 267 326 POTENTIAL.
 FT TRANSMEM 364 410 POTENTIAL.
 FT DOMAIN 723 869 16 X REPEATS, GLY-RICH.

FT REPEAT 723 728 1.
 FT REPEAT 732 737 2.
 FT REPEAT 741 746 3.
 FT REPEAT 750 755 4.
 FT REPEAT 759 764 5.
 FT REPEAT 768 773 6.
 FT REPEAT 777 782 7.
 FT REPEAT 786 791 8.
 FT REPEAT 795 800 9.
 FT REPEAT 806 812 10.
 FT REPEAT 816 821 11.
 FT REPEAT 825 830 12.
 FT REPEAT 834 839 13.
 FT REPEAT 843 848 14.
 FT REPEAT 855 860 15.
 FT REPEAT 864 869 16.
 FT LIPID 563 563 PALMITATE (BY SIMILARITY).
 FT LIPID 569 569 PALMITATE (BY SIMILARITY).
 FT VARIANT 6 6 A -> T (IN STRAIN 2001).
 FT SEQUENCE 1023 AA; 109867 MW; 196D5C0A9A28B54D CRC64;

Query Match 36.8%; Score 7; DB 1; Length 1023;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGNN 7
 |||||
 DB 785 IGGAGNN 791

RESULT 9
 ID PUR6 MYCTU STANDARD; PRT; 174 AA.
 AC P9680;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Phosphoribosylaminoimidazole carboxylase catalytic subunit
 DE (EC 4.1.1.21) (AIR carboxylase) (AIRC).
 GN PURE OR RV3275C OR MT3375 OR MTCY71.15C.
 OS Mycobacterium tuberculosis.
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 CC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean S., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Uferback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: THIS SUBUNIT CAN ALONE TRANSFORM AIR TO CAIR, BUT IN
 CC ASSOCIATION WITH PURK, WHICH POSSESSES AN ATPASE ACTIVITY, AN
 CC ENZYME COMPLEX IS PRODUCED WHICH IS CAPABLE OF CONVERTING AIR TO
 CC CAIR EFFICIENTLY UNDER PHYSIOLOGICAL CONDITION (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: 5-amino-1-(5-phospho-D-ribose)imidazole-4-
 CC carboxylate = 5-amino-1-(5-phospho-D-ribose)imidazole + CO(2).
 CC -!- PATHWAY: De novo purine biosynthesis; sixth step.
 CC -!- SUBUNIT: Homooctamer (By similarity).
 CC -!- SIMILARITY: TO OTHER BACTERIAL PURE, ALSO TO AIR CARBOXYLASE FROM
 CC FUNGI.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z92771; CAB07078.1; -;
 CC EMBL; AE007147; AAK47716.1; -;
 CC PIR; D70979; D70979.
 CC HSP; P09028; 1QZ.
 CC TIGR; MT3375; -;
 CC TuberculList; RV3275C; -;
 CC InterPro; IPR000031; AIR_carboxyl.
 CC Pfam; PF00731; AIRC; 1.
 CC ProDom; PD002193; AIR_carboxyl; 1.
 CC TIGRFAMs; TIGR01162; PurE; 1
 CC Purine biosynthesis; Lyase; Decarboxylase; Complete proteome.

SK SEQUENCE 174 AA; 17674 MW; 18D5F357A7BC8D1 CRC64;
 Query Match 31.6%; Score 6; DB 1; Length 174;
 Best Local Similarity 100.0%; Pred. No. 7.6;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGGAGN 6
 |||||
 DB 121 IGGAGN 126

```
Query Match      31.6%; Score 6; DB 1; Length 351;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


RN [1] SEQUENCE FROM N.A.
 RP STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630 / PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKennay K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Wetman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd.";
 RL Science 269:496-512(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1479;
 RX MEDLINE=97080550; PubMed=8921895;
 RA Sen K., Sikkema D.J., Murphy T.F.;
 RT "Isolation and characterization of the Haemophilus influenzae tolQ,
 tolR, tolA and tolB genes.";
 RL Gene 178:75-81(1996).
 CC -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
 (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -!- SIMILARITY: BELONGS TO THE TOLB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U32722; AAC22040.1; -.
 DR EMBL; U32470; AAC44597.1; -.
 DR PIR; F64064; F64064.
 DR HSSP; P19935; 1CRZ.
 DR TIGR; H10382; -.
 DR Pfam; PF04052; TolB_N; 1.
 KW Transport; Protein transport; Periplasmic; Signal; Complete proteome.
 FT SIGNAL 1 23
 FT CHAIN 24 427
 FT VARIANT 6 6
 FT VARIANT 14 14
 FT VARIANT 17 19
 FT VARIANT 21 21
 FT VARIANT 79 79
 FT VARIANT 129 129
 FT VARIANT 160 160
 FT VARIANT 227 237
 FT VARIANT 322 322
 FT VARIANT 326 326
 FT VARIANT 328 328
 SQ SEQUENCE 427 AA; 44967 MW; 0882201AEE9254B9 CRC64;
 Query Match 31.6%; Score 6; DB 1; Length 427;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 GAGNNT 8
 DB 288 GAGNNT 293
 RESULT 13
 TBG CAEL STANDARD; PRT; 444 AA.
 ID TBG CAEL
 AC P34375;
 DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Tubulin gamma chain (Gamma tubulin).
 GN TBG-1 OR F58A4.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20490776; PubMed=11034903;
 RA Bobinnec Y., Fukuda M., Nishida E.;
 RT "Identification and characterization of Caenorhabditis elegans
 gamma-tubulin in dividing cells and differentiated tissues.";
 RL J. Cell Sci. 113:3747-3759(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=94150748; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Cooper A.,
 RA Bonfield J., Burton J., Connell M., Copesey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
 RA Fulton L., Gardiner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
 RA Sims M., Smaison N., Smith A., Smith M., Sonnhammer E., Staden R.,
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
 RA Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 CC -!- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES.
 CC SUCH AS THE SPINDLE POLES OR THE CENTROSOME. SUGGESTING THAT IT
 IS INVOLVED IN THE MINUS-END NUCLEATION OF MICROTUBULE ASSEMBLY.
 CC -!- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF287259; AAG24513.1; -.
 DR EMBL; Z22179; CAA80164.1; -.
 DR PIR; S40980; S40980.
 DR WormPep; F58A4.8; CE00224.
 DR InterPro; IPR000217; Tubulin.
 DR InterPro; IPR003008; Tubulin_FtsZ.
 DR Pfam; PF00091; tubulin; 1.
 DR Pfam; PF03953; tubulin_C; 1.
 DR PRINTS; PR01161; TUBULIN.
 DR PROSITE; PS00227; TUBULIN; 1.
 KW Microtubules; GTP-binding.
 FT NP_BIND 144 150
 FT GTP (POTENTIAL).
 SQ SEQUENCE 444 AA; 49929 MW; F96CB905B014D6A3 CRC64;
 Query Match 31.6%; Score 6; DB 1; Length 444;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GGAGNN 7
 DB 100 GGAGNN 105
 RESULT 14
 TBG SCHJP STANDARD; PRT; 446 AA.
 ID TBG SCHJP

Search completed: November 21, 2003, 20:59:59
Job time : 4.6075 secs

0

0

GenCore version 5.1.1.6

Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:45:57 ; Search time 23.18 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-85

Perfect score: 19
Sequence: 1 IGGAGNWLHCPTDCRKH 19

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	78.9	2436	12 Q81756	Q81756 hepatitis c
2	15	78.9	3011	12 Q91FES	Q91FES hepatitis c
3	11	57.9	128	12 Q8JYQ3	Q8JYQ3 hepatitis c
4	11	57.9	128	12 Q8JYQ2	Q8JYQ2 hepatitis c
5	11	57.9	128	12 Q8JYQ4	Q8JYQ4 hepatitis c
6	11	57.9	129	12 Q8JYQ1	Q8JYQ1 hepatitis c
7	11	57.9	129	12 Q8JYQ0	Q8JYQ0 hepatitis c
8	11	57.9	137	12 Q914U0	Q914U0 hepatitis c
9	11	57.9	137	12 Q914V3	Q914V3 hepatitis c
10	11	57.9	137	12 Q914U3	Q914U3 hepatitis c
11	11	57.9	137	12 Q914T9	Q914T9 hepatitis c
12	11	57.9	137	12 Q914U4	Q914U4 hepatitis c
13	11	57.9	137	12 Q914V1	Q914V1 hepatitis c
14	11	57.9	137	12 Q914V6	Q914V6 hepatitis c
15	11	57.9	137	12 Q914U6	Q914U6 hepatitis c
16	11	57.9	137	12 Q914V4	Q914V4 hepatitis c

17	11	57.9	137	12	Q914U8	Q914U8 hepatitis c
18	11	57.9	137	12	Q914U1	Q914U1 hepatitis c
19	11	57.9	137	12	Q914U2	Q914U2 hepatitis c
20	11	57.9	137	12	Q914U7	Q914U7 hepatitis c
21	11	57.9	137	12	Q914T6	Q914T6 hepatitis c
22	11	57.9	137	12	Q914T5	Q914T5 hepatitis c
23	11	57.9	137	12	Q914V0	Q914V0 hepatitis c
24	11	57.9	137	12	Q914V5	Q914V5 hepatitis c
25	11	57.9	137	12	Q914U5	Q914U5 hepatitis c
26	11	57.9	137	12	Q914V2	Q914V2 hepatitis c
27	11	57.9	137	12	Q914T7	Q914T7 hepatitis c
28	11	57.9	137	12	Q914U9	Q914U9 hepatitis c
29	11	57.9	640	12	Q68966	Q68966 hepatitis c
30	11	57.9	3011	12	Q03463	Q03463 hepatitis c
31	11	57.9	3011	12	Q9DIT6	Q9DIT6 hepatitis c
32	9	47.4	133	12	Q81497	Q81497 hepatitis c
33	9	47.4	133	12	Q81521	Q81521 hepatitis c
34	9	47.4	133	12	Q81506	Q81506 hepatitis c
35	9	47.4	133	12	Q81539	Q81539 hepatitis c
36	9	47.4	133	12	Q81512	Q81512 hepatitis c
37	9	47.4	133	12	Q81509	Q81509 hepatitis c
38	9	47.4	133	12	Q81515	Q81515 hepatitis c
39	9	47.4	133	12	Q81527	Q81527 hepatitis c
40	9	47.4	133	12	Q81530	Q81530 hepatitis c
41	9	47.4	133	12	Q81533	Q81533 hepatitis c
42	9	47.4	133	12	Q81536	Q81536 hepatitis c
43	9	47.4	133	12	Q81500	Q81500 hepatitis c
44	9	47.4	133	12	Q81524	Q81524 hepatitis c
45	9	47.4	133	12	Q81503	Q81503 hepatitis c

ALIGNMENTS

RESULT 1

Q81756 PRELIMINARY; PRT; 2436 AA.

AC Q81756; 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Genome polyprotein (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=111103;

RN [1]_TaxID=111103;

RP SEQUENCE FROM N.A.

RA Choo Q.-L., Richman K., Han J.;

RT "The nucleotide sequence of the Hepatitis C viral genome.";

RL Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.

DR EMBL; M32084; AAA45677.1; -.

DR HSSP; P27958; 1AIV.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002531; HCV_NSI.

DR InterPro; IPR002518; HCV_NS2.

DR InterPro; IPR004109; HCV_NS3.

DR InterPro; IPR000745; HCV_NS4a.

DR InterPro; IPR001490; HCV_NS4b.

DR InterPro; IPR002868; HCV_NS5a.

DR InterPro; IPR002166; HCV_RGRP.

DR InterPro; IPR001650; Helicase_C.

DR InterPro; IPR007095; RNA_pol_DS_PS.

DR InterPro; IPR007094; RNA_pol_PSVir.

DR Pfam; PF01560; HCV_NSI; 1.

DR Pfam; PF01538; HCV_NS2; 1.

DR Pfam; PF02907; HCV_NS3; 1.

DR Pfam; PF01006; HCV_NS4a; 1.

DR Pfam; PF01001; HCV_NS4b; 1.

DR Pfam; PF01506; HCV_NS5a; 1.

DR Pfam; PF00271; Helicase_C; 1.

DR Pfam; PF00998; Viral_RdRp; 1.

DR ProDom; PD186062; HCV_NSI; 1.

DR EMBL; AF506808; AAM33355.1; -;
 DR InterPro; IPR002531; HCV_NSI.
 DR Pfam; PF01560; HCV_NSI; 1.
 DR ProDom; PD186062; HCV_NSI; 1.
 DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
 DR Polyprotein; Transmembrane.
 FT NON_TER 128 128
 FT NON_TER 128 128
 SQ SEQUENCE 128 AA; 14403 MW; 02CE87238E37F62C CRC64;
 Query Match 57.9%; Score 11; DB 12; Length 128;
 Best Local Similarity 100.0%; Pred. No. 0.00021;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 5 GNNTLHCPTDC 15
 Db 19 GNNTLHCPTDC 29
 RESULT 5
 Q8JYQ4 PRELIMINARY; PRT; 128 AA.
 AC Q8JYQ4;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DE E2 protein (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Quer J., Murillo P., Esteban J.I., Martell M., Esteban R., Guardia J.;
 RT "Sexual transmission of hepatitis C virus from a chronic patient to
 his sexual partner after removal of an intrauterine device.";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF506806; AAM33353.1; -;
 DR InterPro; IPR002531; HCV_NSI.
 DR Pfam; PF01560; HCV_NSI; 1.
 DR ProDom; PD186062; HCV_NSI; 1.
 DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
 DR Polyprotein; Transmembrane.
 FT NON_TER 128 128
 FT NON_TER 128 128
 SQ SEQUENCE 128 AA; 14481 MW; 7EF65920DD3C9ABF CRC64;
 Query Match 57.9%; Score 11; DB 12; Length 128;
 Best Local Similarity 100.0%; Pred. No. 0.00021;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 5 GNNTLHCPTDC 15
 Db 19 GNNTLHCPTDC 29
 RESULT 6
 Q8JYQ1 PRELIMINARY; PRT; 129 AA.
 AC Q8JYQ1;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE E2 protein (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LC4;
 RA Quer J., Murillo P., Esteban J.I., Martell M., Esteban R., Guardia J.;

RT "Sexual transmission of hepatitis C virus from a chronic patient to
 his sexual partner after removal of an intrauterine device.";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF506809; AAM33356.1; -;
 DR InterPro; IPR002531; HCV_NSI.
 DR Pfam; PF01560; HCV_NSI; 1.
 DR ProDom; PD186062; HCV_NSI; 1.
 DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
 DR Polyprotein; Transmembrane.
 FT NON_TER 129 129
 FT NON_TER 129 129
 SQ SEQUENCE 129 AA; 14557 MW; CEB9F0F5A16ADB2B CRC64;
 Query Match 57.9%; Score 11; DB 12; Length 129;
 Best Local Similarity 100.0%; Pred. No. 0.00021;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 5 GNNTLHCPTDC 15
 Db 19 GNNTLHCPTDC 29
 RESULT 7
 Q8JYQ0 PRELIMINARY; PRT; 129 AA.
 AC Q8JYQ0;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE E2 protein (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Quer J., Murillo P., Esteban J.I., Martell M., Esteban R., Guardia J.;
 RT "Sexual transmission of hepatitis C virus from a chronic patient to
 his sexual partner after removal of an intrauterine device.";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF506810; AAM33357.1; -;
 DR InterPro; IPR002531; HCV_NSI.
 DR Pfam; PF01560; HCV_NSI; 1.
 DR ProDom; PD186062; HCV_NSI; 1.
 DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
 DR Polyprotein; Transmembrane.
 FT NON_TER 129 129
 FT NON_TER 129 129
 SQ SEQUENCE 129 AA; 14605 MW; 865657563EC97B7A CRC64;
 Query Match 57.9%; Score 11; DB 12; Length 129;
 Best Local Similarity 100.0%; Pred. No. 0.00021;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 5 GNNTLHCPTDC 15
 Db 19 GNNTLHCPTDC 29
 RESULT 8
 Q914U0 PRELIMINARY; PRT; 137 AA.
 AC Q914U0;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.

```
RC STRAIN=sh;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431847; AAL30697.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI.1.
DR ProDom; PD186062; HCV_NSI.1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 14933 MW; 6730443DE0C31D88 CRC64;

Query Match 57.9%; Score 11; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GNNTLHCPTDC 15
Db 97 GNNTLHCPTDC 107

RESULT 9
Q914V3
ID Q914V3 PRELIMINARY; PRT; 137 AA.
AC Q914V3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=sh;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431834; AAL30684.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI.1.
DR ProDom; PD186062; HCV_NSI.1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 15024 MW; E374842A26F3E2E7 CRC64;

Query Match 57.9%; Score 11; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GNNTLHCPTDC 15
Db 97 GNNTLHCPTDC 107

RESULT 10
Q914U3
ID Q914U3 PRELIMINARY; PRT; 137 AA.
AC Q914U3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
```

```
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=sh;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431844; AAL30694.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI.1.
DR ProDom; PD186062; HCV_NSI.1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 14962 MW; E606A362EF217B6C CRC64;

Query Match 57.9%; Score 11; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GNNTLHCPTDC 15
Db 97 GNNTLHCPTDC 107

RESULT 11
Q914T9
ID Q914T9 PRELIMINARY; PRT; 137 AA.
AC Q914T9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=sh;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431848; AAL30698.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI.1.
DR ProDom; PD186062; HCV_NSI.1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 14977 MW; FEAE343E7FD8E2E7 CRC64;

Query Match 57.9%; Score 11; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GNNTLHCPTDC 15
Db 97 GNNTLHCPTDC 107

RESULT 12
Q914U4
ID Q914U4 PRELIMINARY; PRT; 137 AA.
AC Q914U4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
```

```

OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=sh;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431843; AAL30693.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1 137
FT NON_TER 137 137
SQ SEQUENCE 137 AA; 14990 MW; FEA94432EF33E2E7 CRC64;

Query Match 57.9%; Score 11; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GNNTLHCPTDC 15
Db |||||

RESULT 13
Q914V1 ID Q914V1 PRELIMINARY; PRT; 137 AA.
AC Q914V1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=sh;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431836; AAL30686.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1 137
FT NON_TER 137 137
SQ SEQUENCE 137 AA; 14994 MW; FEA94437FB33E2E7 CRC64;

Query Match 57.9%; Score 11; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GNNTLHCPTDC 15
Db |||||

RESULT 14
Q914V6 ID Q914V6 PRELIMINARY; PRT; 137 AA.
AC Q914V6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)

```

```

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=sh;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431831; AAL30681.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1 137
FT NON_TER 137 137
SQ SEQUENCE 137 AA; 14966 MW; 5EB0443D75188F51 CRC64;

Query Match 57.9%; Score 11; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GNNTLHCPTDC 15
Db |||||

RESULT 15
Q914U6 ID Q914U6 PRELIMINARY; PRT; 137 AA.
AC Q914U6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=sh;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431841; AAL30691.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1 137
FT NON_TER 137 137
SQ SEQUENCE 137 AA; 14991 MW; F403443F8F3E8F87 CRC64;

Query Match 57.9%; Score 11; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GNNTLHCPTDC 15
Db |||||

Search completed: November 21, 2003, 21:08:20
Job time : 23.23 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:22 ; Search time 32.1 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-86

Perfect score: 20

Sequence: 1 TDCFRKHPDATYRCGSGPW 20

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

- 1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	16 AAR84491	Hepatitis C virus
2	20	100.0	20	17 AAR91009	HCV E2 peptide E2
3	20	100.0	20	23 AAO18715	Hepatitis C virus
4	20	100.0	113	16 AAR66632	Hepatitis C virus
5	20	100.0	113	20 AAY01623	Protein encoded by
6	20	100.0	113	20 AAW30596	Hepatitis C virus
7	20	100.0	179	21 AAB18529	Protein encoded by
8	20	100.0	192	19 AAW67009	HCV nucleocapsid c
9	20	100.0	254	22 AAB68043	Amino acid sequenc

10	20	100.0	278	14 AAR33997	Th E2/NS1 protein.
11	20	100.0	363	22 AAB68042	Amino acid sequenc
12	20	100.0	402	14 AAR34438	Sequence of glycop
13	20	100.0	414	14 AAR33589	HCV CKS-NS1S2 fusi
14	20	100.0	414	22 AAB69008	HCV recombinant an
15	20	100.0	454	10 AAR90183	Sequence of hepati
16	20	100.0	454	10 AAR92049	Sequence encoded b
17	20	100.0	454	21 AAB18526	Protein encoded by
18	20	100.0	480	14 AAR33992	HCV-1 E2/NS1 prote
19	20	100.0	531	22 AAR02622	Chimeric HCV E2661
20	20	100.0	621	14 AAR33185	Sequence of subfra
21	20	100.0	622	14 AAR33591	HCV CKS-NS1S1-NS1S
22	20	100.0	622	22 AAB69010	HCV recombinant an
23	20	100.0	637	24 AAB57410	Hepatitis C virus
24	20	100.0	663	17 AAR92935	Hepatitis C virus
25	20	100.0	663	20 AAW67615	Hepatitis C virus
26	20	100.0	738	14 AAR33592	HCV CKS-full lengt
27	20	100.0	738	22 AAB69011	HCV recombinant an
28	20	100.0	2435	13 AAR25135	HCV polypeptide 1.
29	20	100.0	2436	10 AAR92050	Sequence encoded i
30	20	100.0	2436	10 AAR90288	Peptide encoded by
31	20	100.0	2436	13 AAR28582	HCV amino acid seq
32	20	100.0	2772	11 AAR08123	Hepatitis C virus
33	20	100.0	2772	21 AAB18540	Protein encoded by
34	20	100.0	2816	14 AAR34009	HCV-1 polypeptid.
35	20	100.0	2894	13 AAR24440	Composite HCV HC-J
36	20	100.0	2894	16 AAR70230	Composite hepatiti
37	20	100.0	2955	11 AAR08124	Hepatitis C virus
38	20	100.0	2955	20 AAY14975	Amino acid sequenc
39	20	100.0	2955	21 AAB18541	Polypeptid encode
40	20	100.0	3011	13 AAR21519	Compiled HCV sequ
41	20	100.0	3011	14 AAR31621	Hepatitis C virus
42	20	100.0	3011	16 AAR67588	Hepatitis C virus
43	20	100.0	3011	17 AAR90931	Hepatitis C virus
44	20	100.0	3011	18 AAW34480	HCV polypeptid.
45	20	100.0	3011	19 AAW40038	HCV polypeptid.

ALIGNMENTS

RESULT 1

AAR84491
ID AAR84491 standard; peptide; 20 AA.

XX AC AAR84491;

XX DT 06-JAN-1997 (first entry)

XX DT Hepatitis C virus T-cell epitope peptide NS1-23 (residues 583-602).

DE DE Hepatitis C virus; HCV; immunogen; non-structural region;

XX KW Hepatitis C virus; HCV; immunogen; non-structural region;

XX KW immunodominant; T cell epitope; vaccine.

XX OS Hepatitis C virus.

XX OS WO9512677-A2.

PN PN 11-MAY-1995.

PD PD 28-OCT-1994; 94WO-EP03555.

XX PF 28-OCT-1994; 94WO-EP03555.

XX PR 04-NOV-1993; 93EP-0402718.

XX PR (INNO-) INNOGENETICS NV.

XX PA Deleys R, Leroux-Roels G, Maertens G;

XX PI WPI; 1995-193822/25.

XX DR Hepatitis C virus immunogenic polypeptide contg. a T-cell

XX PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in

XX PT production of vaccines, therapeutic agents, etc.

XX PS Claim 26; Page 71; 105pp; English.

XX CC The present sequence is a specifically claimed example of a

XX CC T-cell epitope-containing peptide derived from hepatitis C virus.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 16; Length 20;

Best Local Similarity 100.0%; Pred. No. 8.7e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDCFRKHPDATYRCGSGPW 20
 |||||

Db 1 TDCFRKHPDATYRCGSGPW 20
 |||||

RESULT 2

AAR91009

ID AAR91009 standard; peptide; 20 AA.

XX AC AAR91009;

XX DT 25-SEP-1996 (first entry)

XX DE HCV E2 peptide E2-23 for competition studies.

XX KW HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;

XX KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.

XX OS Synthetic.

XX PN WO9604385-A2.

XX PD 15-FEB-1996.

XX PF 31-JUL-1995; 95WO-EP03031.

XX PR 29-JUL-1994; 94EP-0870132.

XX PA (INNO-) INNOGENETICS NV.

XX PI Bosman F, Buyse M, De Martynoff G, Maertens G;

XX WPI; 1996-129401/13.

XX PT Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope

XX PT proteins - in presence of disulphide bond cleavage agent, to

XX PT produce proteins suitable for direct use in vaccines or diagnostic

XX PT assays of HCV

XX PS Claim 29; Page 67; 146pp; English.

XX CC AAR90965-R90980 and AAR90995-R91015 represent synthetic hepatitis C

XX CC virus (HCV) E1 and E2 peptides used in competition studies. This

XX CC sequence represents a synthetic E2 peptide, and corresponds to residues

XX CC 583-602 of the E2 protein sequence. These sequences are useful for in

XX CC vitro monitoring of HCV disease, or prognosis of the response to

XX CC interferon treatment of patients suffering from HCV infection. These

XX CC sequences compete with the proteins produced by AAT12704-T12709 and

XX CC AAT12961-T12974, which are included in vectors for the production of

XX CC recombinant E1, E2, and E1/E2. The recombinant proteins can then be

XX CC isolated and purified by carrying out a disulphide bond cleavage, or a

XX CC reduction step with a disulphide bond cleavage agent, after lysis of

XX CC recombinant host cells. The constructs containing the purified HCV

XX CC envelope proteins can be used for vaccinating humans against HCV, for in

XX CC vitro detection of HCV antibodies in a sample, and in a serotyping assay

XX CC for detecting one or more serological types of HCV present in a

XX CC biological sample. The constructs can also be immobilised on a solid

XX CC substrate and incorporated into a reversed phase hybridisation assay for

XX CC determining the presence or the genotype of HCV. The new purification

XX CC method preserves the conformation of the recombinantly expressed E1, E2

XX CC and E1/E2, and eliminates contaminating proteins. Antigens isolated

CC using this method are more reactive with human sera than those isolated

CC by known techniques.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 17; Length 20;

Best Local Similarity 100.0%; Pred. No. 8.7e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDCFRKHPDATYRCGSGPW 20
 |||||

Db 1 TDCFRKHPDATYRCGSGPW 20
 |||||

RESULT 3

AAO18715

ID AAO18715 standard; Peptide; 20 AA.

XX AC AAO18715;

XX DT 24-OCT-2002 (first entry)

XX DE Hepatitis C virus E2 protein derived peptide E2-23.

XX KW Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide;

XX KW immunostimulant; vaccine.

XX OS Hepatitis C virus.

XX PN WO200255548-A2.

XX PD 18-JUL-2002.

XX PF 11-JAN-2002; 2002WO-EP00219.

XX PR 11-JAN-2001; 2001US-260699P.

XX PR 30-AUG-2001; 2001US-315768P.

XX PA (INNO-) INNOGENETICS NV.

XX PI Maertens G, Bosman F, Buyse M;

XX WPI; 2002-599657/64.

XX PT New therapeutic vaccine compositions comprising at least one purified

XX PT recombinant hepatitis C virus (HCV) single or specific oligomeric

XX PT recombinant envelope protein E1 or E2, useful for immunizing humans

XX PT from HCV infection

XX PS Claim 4; Page 230; 243pp; English.

XX CC The present invention relates to new therapeutic vaccine compositions for

XX CC inducing hepatitis C virus (HCV)-specific antibodies, comprising a

XX CC composition containing at least one purified recombinant HCV single or

XX CC specific oligomeric recombinant envelope proteins selected from an E1 and

XX CC an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are

XX CC useful for inducing HCV-specific antibodies or for immunising humans

XX CC against HCV. The recombinant HCV E1 and/or E2 proteins are useful as

XX CC vaccines or therapeutics, in HCV screening and confirmatory antibody

XX CC tests, for raising antibodies, in the preparation of medicament, and for

XX CC in vitro monitoring of HCV disease or prognosing the response to

XX CC treatment of patients suffering from HCV infection. The present sequence

XX CC is a peptide derived from the proteins of the invention.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 23; Length 20;

Best Local Similarity 100.0%; Pred. No. 8.7e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDCFRKHPDATYRCGSGPW 20
 |||||

Db 1 TDCFRKHPDATYRCGSGPW 20
 |||||

RESULT 4
 AAR66632
 ID AAR66632 standard; Protein; 113 AA.
 XX AC
 XX AAR66632;
 XX DT 25-MAR-2003 (updated)
 XX DT 31-AUG-1995 (first entry)
 XX Hepatitis C virus J1 NS1 domain consensus protein.
 XX Hepatitis C virus J1 NS1 domain; anti-HCV vaccine development;
 KW non-A non-B virus; diagnostic polypeptides; HCV probes.
 XX OS Hepatitis C virus.
 XX US5372928-A.
 XX PD 13-DEC-1994.
 XX PF 24-FEB-1994; 94US-0201066.
 XX PR 15-SEP-1989; 89US-0408045.
 XX PR 21-DEC-1989; 89US-0456142.
 XX PR 04-JAN-1991; 91US-0637380.
 XX PR 02-AUG-1993; 93US-0101280.
 XX PR 24-FEB-1994; 94US-0201066.
 XX (CHIR) CHIRON CORP.
 PA (NAHE-) NAT INST OF HEALTH JAPAN.
 XX Cha T, Han J, Houghton M, Irvine BD, Kolberg JA;
 PI Miyamura T, Saito I, Weiner AJ;
 XX WPI: 1995-030306/04.
 DR N-PSDB; AAQ79775.
 XX Method of detecting hepatitis C virus polynucleotide - utilises
 PT probe based on DNA of new HCV isolates J1 and J7
 XX Claim 1; Fig 17; 45pp; English.
 XX AAQ79775 encodes AAR66632 the prod. of the hepatitis C virus (HCV)
 CC J1 NS1 domain consensus sequence. They can be used to provide
 CC new oligonucleotides and polypeptides for use in diagnostics,
 CC recombinant protein prodn. and anti-HCV vaccine development.
 CC (Updated on 25-MAR-2003 to correct PF field.)
 XX SQ Sequence 113 AA;
 Query Match 100.0%; Score 20; DB 16; Length 113;
 Best Local Similarity 100.0%; Pred. No. 3.7e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDCFRKHPDATYRCGSGPW 20
 Db 38 TDCFRKHPDATYRCGSGPW 57
 RESULT 5
 AAY01623
 ID AAY01623 standard; Protein; 113 AA.
 XX AC
 XX AAY01623;
 XX DT 21-JUN-1999 (first entry)
 XX Protein encoded by the coding strand of the HCV J1 NS1 domain.
 DE HCV; J7 isolate; J1 isolate; HCV1; immunoassay; asiatic strain;
 KW diagnosis; HCV infection; blood screening; immunisation;

KW antiviral.
 XX Hepatitis C virus.
 XX US5871903-A.
 XX PD 16-FEB-1999.
 XX PF 08-MAY-1995; 95US-0436965.
 XX PR 04-JAN-1991; 91US-0637380.
 XX PR 15-SEP-1989; 89US-0408045.
 XX PR 21-DEC-1989; 89US-0456142.
 XX PR 02-AUG-1993; 93US-0101280.
 XX PR 24-FEB-1994; 94US-0201066.
 XX PR 03-NOV-1994; 94US-0334255.
 XX PR 08-MAY-1995; 95US-0436965.
 XX (CHIR) CHIRON CORP.
 PA (NAHE-) NAT INST OF HEALTH JAPAN.
 XX Miyamura T, Saito I;
 XX WPI: 1999-166619/14.
 DR N-PSDB; AAX26742.
 XX Immunoassays for Asiatic strains of hepatitis C virus - for
 PT diagnosis of infection and screening blood supplies
 XX Disclosure; Fig 17; 43pp; English.
 XX The present sequence is encoded by the consensus sequence of the coding
 CC strand of a new hepatitis C virus (HCV), J1, NS1 domain. The J1 and J7
 CC (also a new HCV isolate) isolates comprise sequences which are distinct
 CC from the prototype HCV isolates, HCV1. The specification describes
 CC immunoassays for HCV based on antigens from Asiatic strains not
 CC cross-reactive with HCV-1. The assays are used for diagnosis of HCV
 CC infection and to screen donated blood. The anti-HCV antibodies are also
 CC useful therapeutically and prophylactically (passive immunisation); in
 CC screening for antiviral agents; for isolation, purification and
 CC identification of non-A, non-B hepatitis virus (e.g. by affinity
 CC chromatography) and to raise anti-idiotypic antibodies (useful for
 CC treatment or diagnosis and to determine immunogenic regions of the
 CC HCV antigens).
 XX SQ Sequence 113 AA;
 Query Match 100.0%; Score 20; DB 20; Length 113;
 Best Local Similarity 100.0%; Pred. No. 3.7e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDCFRKHPDATYRCGSGPW 20
 Db 38 TDCFRKHPDATYRCGSGPW 57
 RESULT 6
 AAW30596
 ID AAW30596 standard; Protein; 113 AA.
 XX AC
 XX AAW30596;
 XX DT 10-MAY-1999 (first entry)
 XX Hepatitis C virus J1 isolate NS1 region.
 DE HCV; non-A, non-B hepatitis; NANBH; diagnosis; vaccine; antigen;
 KW antibody; immunoassay; assay; non-structural domain; NS1.
 XX Hepatitis C virus.
 XX US5856437-A.

```

PD 05-JAN-1999.
XX
XX 03-NOV-1994; 94US-0334255.
XX
XX 04-JAN-1991; 91US-0637380.
PR 15-SEP-1989; 89US-0408045.
PR 21-DEC-1989; 89US-0456142.
PR 02-AUG-1993; 93US-0101280.
PR 24-FEB-1994; 94US-0201066.
PR 03-NOV-1994; 94US-0334255.
XX
XX (CHIR ) CHIRON CORP.
XX (NAME-) NAT INST OF HEALTH JAPAN.
XX
XX Cha T, Han J, Houghton M, Irvine BD, Kolberg JA;
PI Miyamura T, Saito I, Weiner AJ;
XX
XX WPI; 1999-105191/09.
DR N-PSDB; AAX00446.
XX
XX Antigenic polypeptides from J1 and J7 hepatitis C virus isolates -
PT useful as immunoassay reagents, for raising antibodies and as
PT vaccine components
XX
XX Example 5; Fig 17; 44pp; English.
XX
XX This polypeptide comprises a non-structural domain 1 (NS1) region
CC of novel Japanese isolate J1 of hepatitis C virus (HCV), as deduced
CC from an amplified DNA clone (see AAX00446). The sequence shows homology
CC to prototype HCV-1. The invention provides new Japanese isolates,
CC J1 and J7, of HCV. The new isolates have nucleotide and amino acid
CC sequences which are distinct from the prototype HCV-1 isolate.
CC These differences can be exploited for use in diagnostics for
CC NANBH, recombinant protein production and vaccine development.
CC Claimed antigenic polypeptides (see AAX30583-87) can be used: (i) as
CC immunoassay reagents, or standards, to detect HCV antibodies, e.g.
CC for diagnosing infection or screening donated blood; (ii) to
CC generate specific antibodies (used for detecting the corresponding
CC polypeptide, to screen for antiviral agents, for virus isolation
CC and for passive immunisation); (iii) in protective or therapeutic
CC vaccines, and (iv) for isolation of non-A, non-B viruses.
XX
XX Sequence 113 AA;

Query Match 100.0%; Score 20; DB 20; Length 113;
Best Local Similarity 100.0%; Pred. No. 3.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20
Db |||||
38 TDCFRKHPDATYRCGSGPW 57

RESULT 7
AAB18529
ID AAB18529 standard; Protein; 179 AA.
XX
XX AAB18529;
XX
XX 15-JAN-2001 (first entry)
XX
XX Protein encoded by a novel hepatitis C virus cDNA clone 13i.
XX
XX Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
XX viral infectivity; viral replication.
XX
XX Hepatitis C virus.
XX
XX EP1034785-A2.
XX
XX 13-SEP-2000.
XX
XX 16-MAR-1990; 2000EP-0109502.
XX

```

```

XX 17-MAR-1989; 89US-0325338.
PR 20-APR-1989; 89US-0341334.
PR 18-MAY-1989; 89US-0355002.
PR 16-MAR-1990; 90EP-0302866.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Houghton M, Choo Q, Kuo G;
PI
XX WPI; 2000-566891/53.
DR N-PSDB; AAA75285.
XX
XX Novel composition comprising a hepatitis C virus antisense
PT polynucleotide which is complementary to or corresponds to a sense
PT strand of the virus genome, and selectively hybridises to it -
XX
XX Example; Fig 5; 75pp; English.
XX
XX The specification describes a pharmaceutical composition which
CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
CC HCV is characterized by a positive stranded RNA genome which has
CC 40% homology at the polypeptide level to a HCV polyprotein. The
CC antisense polynucleotide binds to cellular polynucleotides which
CC enhance and/or are required for viral infectivity, replicative
CC ability or chronicity. The antisense polynucleotides may also be
CC designed to bind with high specificity, to be of increased stability,
CC to be stable and to have low toxicity. The composition also comprises
CC an agent which causes viral RNA to be inactive. The composition
CC is used for preventing HCV replication in a system. The present
CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
CC course of the invention.
XX
XX Sequence 179 AA;

Query Match 100.0%; Score 20; DB 21; Length 179;
Best Local Similarity 100.0%; Pred. No. 5.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20
Db |||||
73 TDCFRKHPDATYRCGSGPW 92

RESULT 8
AAW67009
ID AAW67009 standard; protein; 192 AA.
XX
XX AAW67009;
XX
XX 02-MAR-1999 (first entry)
XX
XX HCV nucleocapsid core protein.
XX
XX Hepatitis C virus; HCV; nucleocapsid; core protein; analogue; antibody;
XX non-structural protein; thioamide bond; peptide bond.
XX
XX Hepatitis C virus.
XX
XX Key Location/Qualifiers
XX Misc-difference 79 /note= "given in specification as Try"
XX
XX JP10226698-A.
XX
XX 25-AUG-1998.
XX
XX 19-FEB-1997; 97JP-0034702.
XX
XX 19-FEB-1997; 97JP-0034702.
XX
XX (KYOW ) KYOWA MEDEX KK.
XX

```

DR WPI; 1998-515103/44.
 XX Determination of antibody in sample - uses peptide analog absorbed
 PT or chemically bound on carrier as antigen
 XX
 XX
 XX Disclosure; Page 4; 13pp; Japanese.
 PS
 CC This sequence represents the Hepatitis C virus (HCV) nucleocapsid core
 CC protein. The invention relates to peptide analogues derived from HCV
 CC proteins, e.g. AAW67417-W67426, which can be used for the determination
 CC of anti-HCV antibodies in a sample. Preferably the peptide analogues
 CC contain one or more thioamide peptide bonds where at least one oxygen
 CC atom of the peptide bond is replaced by sulphur atom. The peptide
 CC analogues can be adsorbed or chemically bound to a carrier.
 XX
 XX
 SQ Sequence 192 AA;
 Query Match 100.0%; Score 20; DB 19; Length 192;
 Best Local Similarity 100.0%; Pred. No. 5.7e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TDCFRKHPDATYRCGSGPW 20
 Db 133 TDCFRKHPDATYRCGSGPW 152
 RESULT 9
 AAB68043
 ID AAB68043 standard; protein; 254 AA.
 AC AAB68043;
 XX
 XX 29-JUN-2001 (first entry)
 DT
 DE Amino acid sequence of water soluble variant of envelope E2 protein.
 XX
 XX Eo protein; HCV; envelope protein; E2 protein; HCV infection;
 KW HCV attachment.
 XX
 XX Synthetic.
 OS Hepatitis C virus.
 XX
 XX WO200122984-A1.
 PN
 XX 05-APR-2001.
 PD
 XX 26-SEP-2000; 2000WO-US26395.
 PF
 XX 29-SEP-1999; .99US-0407430.
 PR
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 PA
 XX Worman HJ, Mamiya N;
 PI
 XX WPI; 2001-273486/28.
 DR
 XX Treating or preventing hepatitis C virus infection in a subject.
 XX
 PT involves administering hepatitis C virus envelope protein E2 binding
 PT agents -
 XX
 XX Claim 5; Fig 8; 46pp; English.
 PS
 CC The present sequence represents a water soluble variant of a Hepatitis C
 CC virus (HCV) envelope E2 protein. Agents that bind to the HCV E2 protein
 CC (such as the human Eo protein), and so inhibit the attachment of HCV onto
 CC cells (especially liver cells), are used to treat HCV infections in
 CC mammals, in particular humans. The specification also describes a method
 CC for identifying a compound which can be used for treating or preventing
 CC HCV in a subject and which can inhibit the attachment of HCV onto cells
 CC by inhibiting the binding of HCV envelope E2 protein to a cellular
 CC protein associated with HCV attachment and entry into cells. The method
 CC comprises incubating the compound, HCV envelope E2 protein or its variant
 CC and a cellular protein capable of specifically binding to the HCV E2

CC protein under suitable reaction conditions; determining the interactions
 CC between HCV envelope E2 protein and cellular protein in the presence and
 CC absence of the compound; and comparing the interaction to identify a
 CC compound which can inhibit the attachment of HCV onto cells.
 XX
 XX
 SQ Sequence 254 AA;
 Query Match 100.0%; Score 20; DB 22; Length 254;
 Best Local Similarity 100.0%; Pred. No. 7.2e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TDCFRKHPDATYRCGSGPW 20
 Db 178 TDCFRKHPDATYRCGSGPW 197
 RESULT 10
 AAR33997
 ID AAR33997 standard; Protein; 278 AA.
 XX
 AC AAR33997;
 XX
 DT 25-MAR-2003 (updated)
 DT 26-JUL-1993 (first entry)
 XX
 XX Th E2/NS1 protein.
 DE
 XX Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
 KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
 KW domain; immunological; cross-reactive; envelope protein; vaccine;
 KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.
 XX
 XX Synthetic.
 OS
 XX WO9306126-A1.
 PN
 XX 01-APR-1993.
 PD
 XX 11-SEP-1992; 92WO-US07683.
 PF
 XX 13-SEP-1991; 91US-0759575.
 PR
 XX (CHIR) CHIRON CORP.
 PA
 XX Houghton M, Weiner AJ;
 PI
 XX WPI; 1993-117468/14.
 DR
 XX Immuno-reactive hepatitis C virus polypeptide compsns. - contg.
 PT at least 2 sequences from the first variable domain of distinct
 PT HCV isolates
 PT
 XX Disclosure; Fig 3; 106pp; English.
 PS
 XX The sequences given in AAR33992-002 represent a portion of the E2/NS1
 CC protein encoded by group I and group II HCV isolates, from amino acid
 CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
 CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
 CC 30 amino acids which shows large variation between nearly all isolates.
 CC This is an important immunoreactive domain. This putative envelope
 CC glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera
 CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
 CC flaviviruses, both of which confer protective immunity in hosts
 CC vaccinated with these polypeptides. It has been discovered that a
 CC number of important HCV epitopes vary among viral isolates and that
 CC these epitopes can be mapped to specific domains. This meant that
 CC immunologically cross-reactive polypeptides which focus on variable
 CC rather than constant domains can be produced. See also AAQ39134-48
 CC and AAR33982-91.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 XX Sequence 278 AA;
 SQ

```

Query Match      100.0%; Score 20; DB 14; Length 278;
Best Local Similarity 100.0%; Pred. No. 7.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20
DB 214 TDCFRKHPDATYRCGSGPW 233

RESULT 11
AAB68042
ID AAB68042 standard; protein; 363 AA.
XX
AC AAB68042;
XX
DT 29-JUN-2001 (first entry)
XX
DE Amino acid sequence of a Hepatitis C virus envelope E2 protein.
XX
KW E2 protein; HCV; envelope protein; E2 protein; HCV infection;
KW HCV attachment.
XX
OS Hepatitis C virus.
XX
PN WO200122984-A1.
XX
PD 05-APR-2001.
XX
PF 26-SEP-2000; 2000WO-US26395.
XX
PR 29-SEP-1999; 99US-0407430.
XX
PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI Worman HJ, Mamiya N;
XX
DR WPI; 2001-273486/28.
XX
PT Treating or preventing hepatitis C virus infection in a subject,
PT involves administering hepatitis C virus envelope protein E2 binding
PT agents -
XX
PS Claim 3; Fig 7; 46pp; English.
XX
CC The present sequence represents a Hepatitis C virus (HCV) envelope
CC E2 protein. Agents that bind to the HCV E2 protein (such as the human
CC E2 protein), and so inhibit the attachment of HCV onto cells
CC (especially liver cells), are used to treat HCV infections in mammals,
CC in particular humans. The specification also describes a method for
CC identifying a compound which can be used for treating or preventing
CC HCV in a subject and which can inhibit the attachment of HCV onto cells
CC by inhibiting the binding of HCV envelope E2 protein to a cellular
CC protein associated with HCV attachment and entry into cells. The method
CC comprises incubating the compound, HCV envelope E2 protein or its variant
CC and a cellular protein capable of specifically binding to the HCV E2
CC protein under suitable reaction conditions; determining the interactions
CC between HCV envelope E2 protein and cellular protein in the presence and
CC absence of the compound; and comparing the interaction to identify a
CC compound which can inhibit the attachment of HCV onto cells.
XX
SQ Sequence 363 AA;

Query Match      100.0%; Score 20; DB 22; Length 363;
Best Local Similarity 100.0%; Pred. No. 9.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20
DB 200 TDCFRKHPDATYRCGSGPW 219

RESULT 12
AAR34438

```

```

ID AAR34438 standard; Protein; 402 AA.
XX
AC AAR34438;
XX
DT 25-MAR-2003 (updated)
DT 09-AUG-1993 (first entry)
XX
DE Sequence of glycoprotein E2/NS1 in clone HCV1.
XX
KW Hepatitis C virus; envelope protein; glycoprotein; E2/NS1;
KW diagnostic reagent.
XX
OS Hepatitis C virus.
XX
PN EP337626-A1.
XX
PD 21-APR-1993.
XX
PF 08-OCT-1992; 92EP-0117191.
XX
PR 08-OCT-1991; 91JP-0260824.
XX
PA (NAHE-) NAT INST OF HEALTH.
XX
PI Harada S, Honda Y, Miyamura T, Saito I;
XX
DR WPI; 1993-127516/16.
DR N-PSDB; AAQ40330.
XX
PT Diagnostic reagent for hepatitis C virus - comprises second
PT envelope protein or first non-structural protein encoded by HCV
PT gene and has sugar chain
XX
PS Claim 2; Pages 30-32; 58pp; English.
XX
CC Glycoprotein E2/NS1 is derived from the second envelope protein or
CC first non-structural protein encoded by the genome of HCV. The
CC nucleic acid is extracted from the serum of the patient of hepatitis
CC C. The serum is pref. mixed with transfer RNA (tRNA) as a carrier
CC of virus RNA. As a technique of cloning cDNA from the nucleic acid,
CC it is preferred to use polymerase chain reaction method. In the
CC reaction, any commercially available random primers or synthesized
CC DNA having a base sequence similar to that of primer AS1 may be used
CC as a primer. Representative examples of sense primers include S1.
CC (Updated on 25-MAR-2003 to correct FN field.)
XX
SQ Sequence 402 AA;

Query Match      100.0%; Score 20; DB 14; Length 402;
Best Local Similarity 100.0%; Pred. No. 1.1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20
DB 244 TDCFRKHPDATYRCGSGPW 263

RESULT 13
AAR33589
ID AAR33589 standard; protein; 414 AA.
XX
AC AAR33589;
XX
DT 25-MAR-2003 (updated)
DT 05-JUL-1993 (first entry)
XX
DE HCV CKS-NS1S2 fusion antigen.
XX
KW Hepatitis C Virus; non-A, non-B hepatitis virus; NANBH;
KW non-structural protein; CMP-KDO synthetase; CKS fusion protein;
KW CTP: CMP-3-deoxy-manno-octulosonate cytidyl transferase;
KW immunoassay; pHCV-65.
XX

```

OS Hepatitis C Virus.
 PN WO304088-A1.
 XX
 XX
 PD 04-MAR-1993.
 XX
 XX 21-AUG-1992; 92WO-US07188.
 XX
 XX 21-AUG-1991; 91US-0748561.
 PR
 XX (ABBO) ABBOTT LAB.
 PA
 XX
 XX Dailey SH, Desai SM, Devare SG;
 PI
 XX WPI; 1993-093941/11.
 DR
 XX
 XX Hepatitis C assay using recombinant NS1 region antigens - for
 PT detecting antibodies and antigen in body fluids from individuals
 PT exposed to hepatitis C virus
 XX
 XX Claim 2; Page 40-42; 175pp; English.
 PS
 XX Six oligonucleotides representing amino acids 565-731 of the HCV
 CC genome were ligated together and cloned as a 501bp EcoRI/BamHI
 CC fragment into the CKS fusion vector pJO200. The amino acid sequence
 CC of this antigen is designated pHCV-65 (i.e. AAR33589). The resultant
 CC fusion protein HCV CKS-NS1S2 consists of 239 amino acids of CKS,
 CC eight amino acids contributed by linker DNA sequences and 167 amino
 CC acids from the NS1 region of the HCV genome. The fusion protein is
 CC used to detect antibodies and antigen in body fluids from
 CC individuals exposed to HCV.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC
 XX
 XX SQ Sequence 414 AA;
 Query Match 100.0%; Score 20; DB 14; Length 414;
 Best Local Similarity 100.0%; Pred. No. 1.1e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TDCFRKHPDATYRCGSGPW 20
 |||||
 Db 266 TDCFRKHPDATYRCGSGPW 285
 |||||
 RESULT 14
 AAB69008
 ID AAB69008 standard; Peptide; 414 AA.
 AC AAB69008;
 XX
 XX 17-APR-2001 (first entry)
 DT
 XX HCV recombinant antigen pHCV-65 amino acid sequence SEQ ID NO:32.
 DE
 XX Hepatitis C virus; HCV; antigen; detection; antibody.
 KW
 XX Hepatitis C virus.
 OS
 XX US6172189-B1.
 PN
 XX 09-JAN-2001.
 PD
 XX
 XX 02-JUN-1997; 97US-0867611.
 PF
 XX
 XX 19-NOV-1992; 92US-0989843.
 PR
 XX 10-JAN-1994; 94US-0179896.
 PR
 XX 01-MAY-1996; 96US-0646757.
 PR
 XX 24-AUG-1990; 90US-0572822.
 PR
 XX 07-NOV-1990; 90US-0614069.
 PR
 XX 21-AUG-1991; 91US-0748561.
 PR
 XX 21-AUG-1991; 91US-0748566.
 PR
 XX 29-OCT-1991; 91US-0748565.
 XX

PA (ABBO) ABBOTT LAB.
 XX
 XX Devare SG, Desai SM, Casey JM, Dailey SH, Dawson GJ, Gutierrez RA;
 PI Lesniewski RR, Stewart JL, Rupprecht KR;
 XX
 XX WPI; 2001-122352/13.
 DR
 XX
 XX New recombinant antigens representing distinct antigenic regions of
 PT Hepatitis C virus (HCV) genome, useful for detection of antibodies and
 PT antigens in body fluids of individuals exposed to HCV
 XX
 XX Example 11; Column 141-144; 167pp; English.
 PS
 XX The present invention describes recombinant Hepatitis C virus (HCV)
 CC antigens (I). (I) is useful as a reagent for the detection of antibodies
 CC and antigen in body fluids from individuals exposed to HCV. The HCV
 CC assay uses reliable and efficient reagents and methods to accurately
 CC detect the presence of HCV antibodies in samples obtained from
 CC individuals suspected of having HCV infection. AAF32218 to AAF32235,
 CC AAB51371 to AAB51379 and AAB69001 to AAB69032 represent sequences used
 CC in the exemplification of the present invention.
 CC
 XX SQ Sequence 414 AA;
 Query Match 100.0%; Score 20; DB 22; Length 414;
 Best Local Similarity 100.0%; Pred. No. 1.1e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TDCFRKHPDATYRCGSGPW 20
 |||||
 Db 266 TDCFRKHPDATYRCGSGPW 285
 |||||
 RESULT 15
 AAP90183
 ID AAP90183 standard; protein; 454 AA.
 XX
 XX AAP90183;
 AC
 XX 25-MAR-2003 (updated)
 DT 01-NOV-1989 (first entry)
 XX
 XX Sequence of hepatitis C virus cDNA insert in clone k9-1.
 DE
 XX Hepatitis C virus; clone k9-1; probe; vaccine.
 KW
 XX Pan troglodytes.
 OS
 XX
 XX Key Location/Qualifiers
 FH Region 97..454
 FT
 XX GB2212511-A.
 PN
 XX 26-JUL-1989.
 PD
 XX 18-NOV-1988; 88GB-0027024.
 PF
 XX 18-NOV-1987; 87US-0122714.
 PR 30-DEC-1987; 87US-0139886.
 PR 26-FEB-1988; 88US-0161072.
 PR 26-OCT-1988; 88US-0263584.
 PR
 XX (CHIR) CHIRON CORPORATION.
 PA
 XX Houghton M, Choo QL, Kuo G;
 PI
 XX WPI; 1989-215054/30.
 DR N-PSDB; AAN90335.
 XX
 XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
 PT polypeptide(s) and antibodies for diagnosis, prevention and
 PT treatment of infection.
 XX

PS Disclosure; fig 46; 235pp; English.
XX
CC The sequence is the peptide encoded by the hepatitis C virus
CC (HCV) cDNA insert in clone K9-1 (see AAN90335). The polypeptides
CC are used to diagnose HCV-induced NANBH, to raise antibodies for
CC immunoassay or treatment, or to produce vaccines.
CC The region shown overlaps the cDNA of AAN90327.
CC (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ Sequence 454 AA;
Query Match 100.0%; Score 20; DB 10; Length 454;
Best Local Similarity 100.0%; Pred. No. 1.2e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDCFRKHPDATYRCGSGPW 20
| | | | | | | | | | | | | | | | | | | | | |
Db 133 TDCFRKHPDATYRCGSGPW 152
Search completed: November 21, 2003, 20:58:06
Job time : 32.15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.6 Seconds
(without alignments)
79.832 Million cell updates/sec

Title: US-09-973-025-86

Perfect score: 20

Sequence: 1 TDCFRKHPDATYRCGSGPW 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pgp.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pgp.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pgp.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pgp.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pgp.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pgp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-08-612-973-86
2	20	100.0	20	3	US-08-927-597-86
3	20	100.0	20	4	US-08-635-886C-42
4	20	100.0	42	1	US-08-262-037-44
5	20	100.0	68	4	US-08-635-886C-250
6	20	100.0	68	4	US-08-635-886C-262
7	20	100.0	88	1	US-08-440-103-26
8	20	100.0	88	1	US-08-440-103-27
9	20	100.0	88	1	US-08-440-542-26
10	20	100.0	88	1	US-08-440-542-27
11	20	100.0	88	1	US-08-231-368-26
12	20	100.0	88	1	US-08-231-368-27
13	20	100.0	88	1	US-08-440-210-26
14	20	100.0	88	1	US-08-440-210-27
15	20	100.0	88	4	US-09-046-604-26
16	20	100.0	88	4	US-09-046-604-27
17	20	100.0	179	3	US-08-444-818-77
18	20	100.0	278	1	US-08-440-103-15
19	20	100.0	278	1	US-08-440-542-15
20	20	100.0	278	1	US-08-231-368-15
21	20	100.0	278	4	US-08-440-210-15
22	20	100.0	278	4	US-09-046-604-15
23	20	100.0	402	1	US-08-460-806-13
24	20	100.0	402	1	US-08-325-630-13
25	20	100.0	403	2	US-08-483-695-39
26	20	100.0	403	2	US-07-965-285-39
27	20	100.0	403	2	US-08-487-231-39

28	20	100.0	403	3	US-09-201-912-39	Sequence 39, Appl
29	20	100.0	414	1	US-07-748-292-8	Sequence 8, Appl
30	20	100.0	414	3	US-08-867-611-32	Sequence 32, Appl
31	20	100.0	414	5	PCT-US92-06965A-2	Sequence 2, Appl
32	20	100.0	454	3	US-08-444-818-73	Sequence 73, Appl
33	20	100.0	480	1	US-08-440-103-14	Sequence 14, Appl
34	20	100.0	480	1	US-08-440-542-14	Sequence 14, Appl
35	20	100.0	480	1	US-08-231-368-14	Sequence 14, Appl
36	20	100.0	480	1	US-08-440-210-14	Sequence 14, Appl
37	20	100.0	480	4	US-09-046-604-14	Sequence 14, Appl
38	20	100.0	621	1	US-07-748-292-7	Sequence 7, Appl
39	20	100.0	622	3	US-08-867-611-34	Sequence 34, Appl
40	20	100.0	622	5	PCT-US92-06965A-4	Sequence 4, Appl
41	20	100.0	663	3	US-08-824-057-3	Sequence 3, Appl
42	20	100.0	663	4	US-09-415-582-3	Sequence 3, Appl
43	20	100.0	663	4	US-09-693-596-4	Sequence 4, Appl
44	20	100.0	738	3	US-08-867-611-35	Sequence 35, Appl
45	20	100.0	738	5	PCT-US92-06965A-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-612-973-86
; Sequence 86, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-612-973-86

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 3,1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20
|||||

```
Db      1  TDCFRKHPDATYSRCGSPW 20

RESULT 2
US-08-927-597-86
; Sequence 86, Application US/08927597
; Patent No. 6245503
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-927-597-86

Query Match      100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.le-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  TDCFRKHPDATYSRCGSPW 20
      |||||
Db      1  TDCFRKHPDATYSRCGSPW 20

RESULT 3
US-08-635-886C-42
; Sequence 42, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555

; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 42
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-42

Query Match      100.0%; Score 20; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.le-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  TDCFRKHPDATYSRCGSPW 20
      |||||
Db      1  TDCFRKHPDATYSRCGSPW 20

RESULT 4
US-08-262-037-44
; Sequence 44, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-759-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-262-037-44
```

Query Match 100.0%; Score 20; DB 1; Length 42;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDCFRKHPDATYRCGSGPW 20
Db 5 TDCFRKHPDATYRCGSGPW 24

RESULT 5
US-08-635-886C-250
; Sequence 250, Application US/08635886C
; Patent No. 655114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635.886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 250
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-250

Query Match 100.0%; Score 20; DB 4; Length 68;
Best Local Similarity 100.0%; Pred. No. 8.4e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDCFRKHPDATYRCGSGPW 20
Db 13 TDCFRKHPDATYRCGSGPW 32

RESULT 6
US-08-635-886C-262
; Sequence 262, Application US/08635886C
; Patent No. 655114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635.886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 262
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-262

Query Match 100.0%; Score 20; DB 4; Length 68;
Best Local Similarity 100.0%; Pred. No. 8.4e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDCFRKHPDATYRCGSGPW 20

Db 13 TDCFRKHPDATYRCGSGPW 32

RESULT 7
US-08-440-103-26
; Sequence 26, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440.103
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231.368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-103-26

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDCFRKHPDATYRCGSGPW 20
Db 30 TDCFRKHPDATYRCGSGPW 49

RESULT 8
US-08-440-103-27
; Sequence 27, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA

```

;
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,103
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-103-27

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYSRCGSGPW 20
Db 30 TDCFRKHPDATYSRCGSGPW 49

RESULT 9
US-08-440-542-26
; Sequence 26, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,542
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-542-27

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYSRCGSGPW 20
Db 30 TDCFRKHPDATYSRCGSGPW 49

RESULT 9
US-08-440-542-26
; Sequence 26, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,542
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.

```

```

;
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-542-26

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYSRCGSGPW 20
Db 30 TDCFRKHPDATYSRCGSGPW 49

RESULT 10
US-08-440-542-27
; Sequence 27, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,542
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-542-27

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 TDCFRKHPDATYRCGSGPW 20
Db 30 TDCFRKHPDATYRCGSGPW 49

RESULT 11

US-08-231-368-26
; Sequence 26, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-231-368-26

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDCFRKHPDATYRCGSGPW 20
Db 30 TDCFRKHPDATYRCGSGPW 49

RESULT 12

US-08-231-368-27
; Sequence 27, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA

; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-231-368-27

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDCFRKHPDATYRCGSGPW 20
Db 30 TDCFRKHPDATYRCGSGPW 49

RESULT 13

US-08-440-210-26
; Sequence 26, Application US/08440210
; Patent No. 5766845
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,210
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-210-26

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 1e-13; 0; Indels 0;
Matches 20; Conservative 0; Mismatches 0; Gaps 0;

Qy 1 TDCFRKHPDATYRCGSGPW 20
|||||
Db 30 TDCFRKHPDATYRCGSGPW 49

RESULT 14
US-08-440-210-27
Sequence 27, Application US/08440210
Patent No. 5766845
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,210
FILING DATE: 12-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-210-27

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 1e-13; 0; Indels 0;
Matches 20; Conservative 0; Mismatches 0; Gaps 0;

Qy 1 TDCFRKHPDATYRCGSGPW 20
|||||
Db 30 TDCFRKHPDATYRCGSGPW 49

RESULT 15
US-09-046-604-26
Sequence 26, Application US/09046604
Patent No. 6303292
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/046,604
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-046-604-26

Query Match 100.0%; Score 20; DB 4; Length 88;
Best Local Similarity 100.0%; Pred. No. 1e-13; 0; Indels 0;
Matches 20; Conservative 0; Mismatches 0; Gaps 0;

Qy 1 TDCFRKHPDATYRCGSGPW 20
|||||
Db 30 TDCFRKHPDATYRCGSGPW 49

Search completed: November 21, 2003, 21:15:17
Job time : 10.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time 19.5 Seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-86

Perfect score: 20

Sequence: 1 TDCFRKHPDATYRCGSGPW 20

Scoring table:

OLIGO Gapex 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PTCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PTCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	10	US-09-973-025-86
2	20	100.0	20	11	US-09-973-025-86
3	20	100.0	20	11	US-09-973-025-86
4	20	100.0	20	11	US-09-973-025-86
5	20	100.0	20	12	US-09-973-025-86
6	20	100.0	254	10	US-09-407-430-3
7	20	100.0	363	10	US-09-407-430-2
8	20	100.0	637	12	US-10-187-257-4
9	20	100.0	637	12	US-10-265-083-2
10	20	100.0	2894	10	US-09-941-611-23
11	20	100.0	2894	10	US-10-044-995-23
12	20	100.0	3011	9	US-09-916-359-2
13	20	100.0	3011	16	US-10-232-643-6
14	11	55.0	20	15	US-10-127-746-17
15	11	55.0	20	15	US-10-150-165-17

16	11	55.0	176	10	US-09-921-397-81	Sequence 81, Appl
17	11	55.0	250	10	US-09-952-572-8	Sequence 8, Appl
18	11	55.0	363	12	US-10-128-587A-97	Sequence 97, Appl
19	11	55.0	363	15	US-10-128-590-97	Sequence 97, Appl
20	11	55.0	3011	9	US-09-742-659-4	Sequence 4, Appl
21	11	55.0	3011	10	US-09-238-076-20	Sequence 20, Appl
22	11	55.0	3011	10	US-09-952-572-9	Sequence 9, Appl
23	11	55.0	3011	10	US-09-747-419-20	Sequence 20, Appl
24	11	55.0	3011	11	US-09-891-894-3	Sequence 3, Appl
25	11	55.0	3011	11	US-09-995-937-20	Sequence 20, Appl
26	11	55.0	3011	11	US-09-917-563-20	Sequence 20, Appl
27	11	55.0	3011	12	US-10-184-150-3	Sequence 3, Appl
28	11	55.0	3011	15	US-10-259-275-20	Sequence 20, Appl
29	11	55.0	3012	10	US-09-238-076-2	Sequence 2, Appl
30	11	55.0	3012	11	US-09-995-937-2	Sequence 2, Appl
31	11	55.0	3012	11	US-09-917-563-2	Sequence 4, Appl
32	10	50.0	350	10	US-09-929-955-4	Sequence 4, Appl
33	10	50.0	350	14	US-10-104-966-4	Sequence 4, Appl
34	10	50.0	3011	10	US-09-529-955-1	Sequence 1, Appl
35	10	50.0	3011	14	US-10-104-966-1	Sequence 1, Appl
36	9	45.0	2940	12	US-10-226-629A-13	Sequence 13, Appl
37	8	40.0	20	10	US-09-973-025-87	Sequence 87, Appl
38	8	40.0	20	11	US-09-899-303-87	Sequence 87, Appl
39	8	40.0	20	11	US-09-995-808-87	Sequence 87, Appl
40	8	40.0	20	11	US-09-995-860-87	Sequence 87, Appl
41	8	40.0	20	12	US-09-995-791-87	Sequence 87, Appl
42	8	40.0	30	16	US-10-318-200-15	Sequence 15, Appl
43	8	40.0	45	16	US-10-318-200-16	Sequence 16, Appl
44	8	40.0	290	12	US-10-128-587A-3	Sequence 3, Appl
45	8	40.0	290	15	US-10-128-590-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1

US-09-973-025-86
; Sequence 86, Application US/09973025
; Publication No. US20020182706A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973.025
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100

; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 86:
US-09-973-025-86

Query Match 100.0%; Score 20; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 TDCFRKHPDATYRCGSGPW 20

RESULT 2

US-09-899-303-86
; Sequence 86, Application US/09899303
; Publication No. US20030036110A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA: US/09/899,303

APPLICATION NUMBER: US/09/899,303

FILING DATE: 06-Jul-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 86:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 86:

US-09-899-303-86

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20
| | | | | | | | | | | | | | | | | | | | | |

Db 1 TDCFRKHPDATYRCGSGPW 20

RESULT 3

US-09-995-808-86
; Sequence 86, Application US/09995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:

; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808

CURRENT FILING DATE: 2001-11-29

NUMBER OF SEQ ID NOS: 122

SOFTWARE: Patent In 3.1

SEQ ID NO 86

LENGTH: 20

TYPE: PRT

ORGANISM: Hepatitis C virus

US-09-995-808-86

Query Match 100.0%; Score 20; DB 11; Length 20;

Best Local Similarity 100.0%; Pred. No. 3.2e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20

| | | | | | | | | | | | | | | | | | | | | |

Db 1 TDCFRKHPDATYRCGSGPW 20

RESULT 4

US-09-995-860-86
; Sequence 86, Application US/09995860
; Publication No. US20030118603A1
; GENERAL INFORMATION:

; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and

; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860

CURRENT FILING DATE: 2001-11-29

NUMBER OF SEQ ID NOS: 122

SOFTWARE: Patent In 3.1

SEQ ID NO 86

LENGTH: 20

TYPE: PRT

ORGANISM: Hepatitis C virus

US-09-995-860-86

Query Match 100.0%; Score 20; DB 11; Length 20;

Best Local Similarity 100.0%; Pred. No. 3.2e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20

| | | | | | | | | | | | | | | | | | | | | |

Db 1 TDCFRKHPDATYRCGSGPW 20

RESULT 5

US-09-995-791-86
; Sequence 86, Application US/09995791
; Publication No. US20030147918A1
; GENERAL INFORMATION:

; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791

CURRENT FILING DATE: 2001-11-29

NUMBER OF SEQ ID NOS: 122

SOFTWARE: Patent In 3.1

SEQ ID NO 86

LENGTH: 20

TYPE: PRT

ORGANISM: Hepatitis C virus

US-09-995-791-86

Query Match 100.0%; Score 20; DB 11; Length 20;

Best Local Similarity 100.0%; Pred. No. 3.2e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20

| | | | | | | | | | | | | | | | | | | | | |

Db 1 TDCFRKHPDATYRCGSGPW 20


```
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-86

Query Match      100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYSRGSGPW 20
   |||||
Db 1 TDCFRKHPDATYSRGSGPW 20
   |||||

RESULT 6
US-09-407-430-3
; Sequence 3, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-407-430-3

Query Match      100.0%; Score 20; DB 10; Length 254;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYSRGSGPW 20
   |||||
Db 178 TDCFRKHPDATYSRGSGPW 197
   |||||

RESULT 7
US-09-407-430-2
; Sequence 2, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-09-407-430-2

Query Match      100.0%; Score 20; DB 10; Length 363;
Best Local Similarity 100.0%; Pred. No. 3.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYSRGSGPW 20
   |||||
Db 200 TDCFRKHPDATYSRGSGPW 219
   |||||

RESULT 8
US-10-187-257-4
; Sequence 4, Application US/10187257
; Publication No. US20030138458A1
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: CORTES, Steve
; APPLICANT: O'HAGAN, Derek
; TITLE OF INVENTION: HCV E1E2 VACCINE COMPOSITIONS
; FILE REFERENCE: 2302-17206
; CURRENT APPLICATION NUMBER: US/10/187,257
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
; OTHER INFORMATION: amino acid
US-10-187-257-4

Query Match      100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 4.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYSRGSGPW 20
   |||||
Db 411 TDCFRKHPDATYSRGSGPW 430
   |||||

RESULT 9
US-10-265-083-2
; Sequence 2, Application US/10265083
; Publication No. US20030170273A1
; GENERAL INFORMATION:
; APPLICANT: O'HAGAN, Derek
; APPLICANT: VALIANTE, Nicholas
; TITLE OF INVENTION: ADJUVANT COMPOSITIONS
; FILE REFERENCE: 2302-18357.30
; CURRENT APPLICATION NUMBER: US/10/265,083
; CURRENT FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7
; OTHER INFORMATION: region
US-10-265-083-2

Query Match      100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 4.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYSRGSGPW 20
   |||||
Db 411 TDCFRKHPDATYSRGSGPW 430
   |||||

RESULT 10
US-09-941-611-23
; Sequence 23, Application US/09941611
; Patent No. US20020106640A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSWUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; ANTIBODIES TO HEPATITIS C VIRUS
```

NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/941,611
FILING DATE: 30-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: 1995-02-21
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-941-611-23
Query Match 100.0%; Score 20; DB 10; Length 2894;
Best Local Similarity 100.0%; Pred. No. 1.6e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDCFRKHPDATYRCGSGPW 20
Db 583 TDCFRKHPDATYRCGSGPW 602
RESULT 11
US-10-044-995-23
Sequence 23, Application US/10044995
Publication No. US20030049685A1
GENERAL INFORMATION:
APPLICANT: DELEYS, ROBERT J
POLLET, DIRK
MAERTENS, GEERT
VAN HEUVERSWIN, HUGO
TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
ANTIBODIES TO HEPATITIS C VIRUS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/044,995
FILING DATE: 15-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: <Unknown>
APPLICATION NUMBER: US 07/920,286
FILING DATE: 14-OCT-1992
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-044-995-23
Query Match 100.0%; Score 20; DB 15; Length 2894;
Best Local Similarity 100.0%; Pred. No. 1.6e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDCFRKHPDATYRCGSGPW 20
Db 583 TDCFRKHPDATYRCGSGPW 602
RESULT 12
US-09-916-359-2
Sequence 2, Application US/09916359
Patent No. US20020034734A1
GENERAL INFORMATION:
APPLICANT: Veronique Barban
TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
TREATING C HEPATITIS
FILE REFERENCE: PMCF97-03A
CURRENT APPLICATION NUMBER: US/09/916,359
CURRENT FILING DATE: 2001-07-26
PRIOR APPLICATION NUMBER: 09/388,874
PRIOR FILING DATE: 1999-09-02
PRIOR APPLICATION NUMBER: 97/02,887
PRIOR FILING DATE: 1997-03-06
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 3011
TYPE: PRT
ORGANISM: Virus
US-09-916-359-2
Query Match 100.0%; Score 20; DB 9; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.6e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDCFRKHPDATYRCGSGPW 20
Db 583 TDCFRKHPDATYRCGSGPW 602

Db 583 TDCFRKHPDATYRCGSGPW 602

RESULT 13

US-10-232-643-6

Sequence 6, Application US/10232643

Publication No. US20030129586A1

GENERAL INFORMATION:

APPLICANT: HOUGHTON, MICHAEL

CHOO, QUI-LIM

HAN, JANG

CHOE, JOONHO

TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING

HELICASE ACTIVITY AND IMPROVED SOLUBILITY

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: CHIRON CORPORATION

STREET: Intellectual Property - R440, P.O. Box 8097

CITY: Emeryville

STATE: California

COUNTRY: U.S.A.

ZIP: 94662-8097

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/232,643

FILING DATE: 30-Aug-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/483,799

FILING DATE: <Unknown>

APPLICATION NUMBER: 08/529,169

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Harbin, Alisa A.

REGISTRATION NUMBER: 33,895

REFERENCE/DOCKET NUMBER: 0100.005

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 923-3274

TELEFAX: (510) 655-3542

TELEX: n/a

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 3011 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Duplication

LOCATION: 9

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Lys or Arg"

FEATURE:

NAME/KEY: Duplication

LOCATION: 11

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Asn or Thr"

FEATURE:

NAME/KEY: Duplication

LOCATION: 176

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Ile or Thr"

FEATURE:

NAME/KEY: Duplication

LOCATION: 334

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Met or Val"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2502

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Leu or Phe"

LOCATION: 603

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Leu or Ile"

FEATURE:

NAME/KEY: Duplication

LOCATION: 848

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Tyr or Asn"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1114

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Pro or Ser"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1117

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Ser or Thr"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1276

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Pro or Leu"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1454

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Cys or Tyr"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1471

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Thr or Ser"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1877

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Glu or Gly"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1948

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Leu or His"

FEATURE:

NAME/KEY: Duplication

LOCATION: 1949

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Ser or Cys"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2021

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Gly or Val"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2349

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Thr or Ser"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2385

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Tyr or Phe"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2386

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Ser or Ala"

FEATURE:

NAME/KEY: Duplication

LOCATION: 2502

OTHER INFORMATION: /note= "There exists a heterogeneity at this position - Xaa = Leu or Phe"

```
;
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2690
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Arg or Gly"
;
; NAME/KEY: Duplication
; LOCATION: 2921
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Arg or Gly"
;
; NAME/KEY: Duplication
; LOCATION: 2996
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Leu or Pro"
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-232-643-6
```

```
Query Match 100.0%; Score 20; DB 16; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.6e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYSRGSGPW 20
| | | | | | | | | | | | | | | | | | | | | |
DB 583 TDCFRKHPDATYSRGSGPW 602
```

```
RESULT 14
US-10-127-746-17
; Sequence 17, Application US/10127746
; Publication No. US20030091590A1
; GENERAL INFORMATION:
; APPLICANT: Schnell, Matthias J.
; APPLICANT: Pomerantz, Roger J.
; TITLE OF INVENTION: Recombinant Rhabdoviruses as Live Viral
; TITLE OF INVENTION: Vaccines
; FILE REFERENCE: SCH02.CP201
; CURRENT APPLICATION NUMBER: US/10/127,746
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 09/494,262
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 09/761,312
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/285,552
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic peptide
US-10-127-746-17
```

```
Query Match 55.0%; Score 11; DB 15; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.7e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 ATYSRCGSGPW 20
| | | | | | | | | | | |
DB 2 ATYSRCGSGPW 12
```

```
RESULT 15
US-10-150-165-17
; Sequence 17, Application US/10150165
; Publication No. US20030124146A1
; GENERAL INFORMATION:
; APPLICANT: Schnell, Matthias J.
; APPLICANT: Pomerantz, Roger J.
; TITLE OF INVENTION: Recombinant Rhabdoviruses as Live Viral
; TITLE OF INVENTION: Vaccines
```

```
; FILE REFERENCE: SCH02.PC301
; CURRENT APPLICATION NUMBER: US/10/150,165
; CURRENT FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 09/494,262
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 09/761,312
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: SCH02-CP201
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 60/285,552
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic peptide
US-10-150-165-17

Query Match 55.0%; Score 11; DB 15; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.7e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 ATYSRCGSGPW 20
| | | | | | | | | | | |
DB 2 ATYSRCGSGPW 12

Search completed: November 21, 2003, 22:19:40
Job time : 20.5 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 163.1 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-86

Perfect score: 20
Sequence: 1 TDCPRKHPDATYRCGSGPW 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA_Main.*
1: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US097A_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US097B_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US099A_COMB.pep.*
25: /cgn2_6/ptodata/1/paa/US099B_COMB.pep.*
26: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
27: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
28: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
29: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
30: /cgn2_6/ptodata/1/paa/US104_COMB.pep.*
31: /cgn2_6/ptodata/1/paa/US106_COMB.pep.*
32: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	ID	Description
1	20	100.0	20	13	US-08-974-685-42	Sequence 42, Appl

2	20	100.0	20	13	US-08-974-690-42	Sequence 42, Appl
3	20	100.0	20	13	US-08-974-690A-42	Sequence 42, Appl
4	20	100.0	20	13	US-08-974-690B-42	Sequence 42, Appl
5	20	100.0	20	13	US-08-974-690C-42	Sequence 42, Appl
6	20	100.0	20	23	US-09-899-303-86	Sequence 86, Appl
7	20	100.0	20	23	US-09-899-303A-86	Sequence 86, Appl
8	20	100.0	20	25	US-09-973-025-86	Sequence 86, Appl
9	20	100.0	20	25	US-09-995-791-86	Sequence 86, Appl
10	20	100.0	20	25	US-09-995-808-86	Sequence 86, Appl
11	20	100.0	20	25	US-09-995-860-86	Sequence 86, Appl
12	20	100.0	20	26	US-10-020-510-86	Sequence 86, Appl
13	20	100.0	20	29	US-10-321-798-86	Sequence 86, Appl
14	20	100.0	42	8	US-08-475-482-44	Sequence 44, Appl
15	20	100.0	42	8	US-08-477-072-44	Sequence 44, Appl
16	20	100.0	42	8	US-08-477-582-44	Sequence 44, Appl
17	20	100.0	42	8	US-08-480-253-44	Sequence 44, Appl
18	20	100.0	68	13	US-08-974-690C-250	Sequence 250, App
19	20	100.0	68	13	US-08-974-690C-262	Sequence 262, App
20	20	100.0	88	8	US-08-471-498-26	Sequence 26, Appl
21	20	100.0	88	8	US-08-471-498-27	Sequence 27, Appl
22	20	100.0	112	8	US-08-436-966-18	Sequence 18, Appl
23	20	100.0	112	8	US-08-436-966-19	Sequence 19, Appl
24	20	100.0	179	8	US-08-403-590B-77	Sequence 77, Appl
25	20	100.0	179	8	US-08-444-112-77	Sequence 77, Appl
26	20	100.0	254	1	PCT-US00-26395-3	Sequence 3, Appl
27	20	100.0	254	18	US-09-407-430-3	Sequence 3, Appl
28	20	100.0	278	8	US-08-471-498-15	Sequence 15, Appl
29	20	100.0	333	30	US-10-445-724-3	Sequence 3, Appl
30	20	100.0	333	32	US-60-409-909-4	Sequence 4, Appl
31	20	100.0	363	1	PCT-US00-26395-2	Sequence 2, Appl
32	20	100.0	363	18	US-09-407-430-2	Sequence 2, Appl
33	20	100.0	414	3	US-07-748-561-2	Sequence 32, Appl
34	20	100.0	414	3	US-07-989-843-32	Sequence 32, Appl
35	20	100.0	414	8	US-08-463-849-32	Sequence 32, Appl
36	20	100.0	414	8	US-08-463-884-32	Sequence 32, Appl
37	20	100.0	414	20	US-09-690-359-32	Sequence 32, Appl
38	20	100.0	454	8	US-08-403-590B-73	Sequence 73, Appl
39	20	100.0	454	8	US-08-444-112-73	Sequence 14, Appl
40	20	100.0	480	8	US-08-471-498-14	Sequence 7, Appl
41	20	100.0	531	21	US-09-721-480-7	Sequence 4, Appl
42	20	100.0	622	3	US-07-748-561-4	Sequence 34, Appl
43	20	100.0	622	3	US-07-989-843-34	Sequence 34, Appl
44	20	100.0	622	8	US-08-463-849-34	Sequence 34, Appl
45	20	100.0	622	8	US-08-463-884-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1

US-08-974-685-42
; Sequence 42, Application US/08974685
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
; THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 181
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,685
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 42:

US-08-974-685-42

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20
DB 1 TDCFRKHPDATYRCGSGPW 20

RESULT 2

US-08-974-690-42
Sequence 42, Application US/08974690
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
DELEYS, ROBERT
MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 166
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHUYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/635,886
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

US-08-974-690-42

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20
DB 1 TDCFRKHPDATYRCGSGPW 20

RESULT 3

US-08-974-690A-42
Sequence 42, Application US/08974690A
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
DELEYS, ROBERT
MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHUYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690A
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 42:

US-08-974-690A-42

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20
DB 1 TDCFRKHPDATYRCGSGPW 20

RESULT 4

US-08-974-690B-42
Sequence 42, Application US/08974690B
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
DELEYS, ROBERT
MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION NUMBER: US/08/974,690B
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-08-974-690B-42

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20
|||||
DB 1 TDCFRKHPDATYRCGSGPW 20

RESULT 5
US-08-974-690C-42
; Sequence 42, Application US/08574690C
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/08/974,690C
; CURRENT FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-974-690C-42

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20

DB 1 TDCFRKHPDATYRCGSGPW 20
|||||

RESULT 6
US-09-899-303-86
; Sequence 86, Application US/09899303
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 86:
US-09-899-303-86

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20
|||||
DB 1 TDCFRKHPDATYRCGSGPW 20

RESULT 7
US-09-899-303A-86
; Sequence 86, Application US/09899303A
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.

STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303A
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 86:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 86:
US-09-899-303A-86

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYSRCGSPW 20
|||||
DB 1 TDCFRKHPDATYSRCGSPW 20
|||||

RESULT 8
US-09-973-025-86
; Sequence 86, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 86:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 86:
US-09-973-025-86

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYSRCGSPW 20
|||||
DB 1 TDCFRKHPDATYSRCGSPW 20
|||||

RESULT 9
US-09-995-791-86
; Sequence 86, Application US/09995791
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 86
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-86

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYSRCGSPW 20
|||||
DB 1 TDCFRKHPDATYSRCGSPW 20
|||||

RESULT 10
US-09-995-808-86
; Sequence 86, Application US/09995808
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 86
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-86

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;


```
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TDCFRKHPDATYRCGSGPW 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 TDCFRKHPDATYRCGSGPW 20

RESULT 11
US-09-995-860-86
; Sequence 86, Application US/09995860
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 86
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-86

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TDCFRKHPDATYRCGSGPW 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 TDCFRKHPDATYRCGSGPW 20

RESULT 12
US-10-020-510-86
; Sequence 86, Application US/10020510
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-72
; CURRENT APPLICATION NUMBER: US/10/020,510
; CURRENT FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 86
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-020-510-86

Query Match 100.0%; Score 20; DB 26; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TDCFRKHPDATYRCGSGPW 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 TDCFRKHPDATYRCGSGPW 20

RESULT 13
US-10-321-798-86
; Sequence 86, Application US/10321798
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-93
; CURRENT APPLICATION NUMBER: US/10/321,798
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: 60/418,358
; PRIOR FILING DATE: 2002-10-16
```

```
; PRIOR APPLICATION NUMBER: 10/020,510
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 86
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-321-798-86

Query Match 100.0%; Score 20; DB 29; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TDCFRKHPDATYRCGSGPW 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 TDCFRKHPDATYRCGSGPW 20

RESULT 14
US-08-475-482-44
; Sequence 44, Application US/08475482
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,482
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 17-June-1994
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-475-482-44
```

Job time : 163.1 secs

Query Match 100.0%; Score 20; DB 8; Length 42;
Best Local Similarity 100.0%; Pred. No. 1e-13; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20
|||||
DB 5 TDCFRKHPDATYRCGSGPW 24

RESULT 15
US-08-477-072-44
; Sequence 44, Application US/08477072
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,072
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 17-June-1994
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
; US-08-477-072-44

Query Match 100.0%; Score 20; DB 8; Length 42;
Best Local Similarity 100.0%; Pred. No. 1e-13; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20
|||||
DB 5 TDCFRKHPDATYRCGSGPW 24

Search completed: November 21, 2003, 22:09:53

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.55 Seconds
(without alignments)
117.536 Million cell updates/sec

Title: US-09-973-025-86
Perfect score: 20
Sequence: 1 TDCFRKHPDATYRCGSGPW 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents AA New:
1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	20	100.0	20	US-10-651-165-42
2	20	100.0	68	US-10-651-165-250
3	20	100.0	68	US-10-651-165-262
4	20	100.0	333	1 PCT-US03-19834-3
5	20	100.0	333	US-10-655-562-4
6	20	100.0	637	1 PCT-US03-33610-4
7	20	100.0	3011	1 PCT-US03-19834-2
8	12	60.0	70	6 US-10-651-165-264
9	11	55.0	30	6 US-10-296-734-486
10	11	55.0	68	6 US-10-651-165-252
11	11	55.0	1997	6 US-10-296-734-816
12	11	55.0	3011	6 US-10-296-734-406
13	11	55.0	5985	6 US-10-296-734-810
14	10	50.0	68	6 US-10-651-165-251
15	9	45.0	70	6 US-10-651-165-263
16	9	45.0	3033	6 US-10-009-002-5
17	8	40.0	9	6 US-10-651-165-171
18	8	40.0	20	6 US-10-651-165-41
19	8	40.0	20	6 US-10-651-165-43
20	8	40.0	30	6 US-10-296-734-484
21	8	40.0	30	6 US-10-685-435-26
22	8	40.0	45	6 US-10-685-435-27
23	8	40.0	68	6 US-10-651-165-253
24	8	40.0	68	6 US-10-651-165-254
25	8	40.0	68	6 US-10-651-165-255
26	8	40.0	68	6 US-10-651-165-256

27	8	40.0	68	6	US-10-651-165-257	Sequence 257, App
28	8	40.0	68	6	US-10-651-165-258	Sequence 258, App
29	8	40.0	68	6	US-10-651-165-259	Sequence 259, App
30	8	40.0	68	6	US-10-651-165-260	Sequence 260, App
31	8	40.0	68	6	US-10-651-165-261	Sequence 261, App
32	8	40.0	347	6	US-10-664-391-9	Sequence 9, Appli
33	8	40.0	539	6	US-10-664-391-11	Sequence 11, Appli
34	8	40.0	1026	1	PCT-US03-20409-3	Sequence 3, Appli
35	8	40.0	2010	6	US-10-296-734-814	Sequence 814, App
36	8	40.0	2280	1	PCT-US03-20322-211	Sequence 211, App
37	7	35.0	33	6	US-10-685-435-28	Sequence 28, Appli
38	6	30.0	9	1	PCT-US03-31303-125	Sequence 125, App
39	6	30.0	9	6	US-10-677-754-125	Sequence 125, App
40	6	30.0	20	6	US-10-651-165-131	Sequence 131, App
41	6	30.0	22	6	US-10-651-165-130	Sequence 130, App
42	6	30.0	70	6	US-10-651-165-108	Sequence 108, App
43	6	30.0	268	6	US-10-425-114A-38848	Sequence 38848, A
44	6	30.0	290	6	US-10-425-114A-44768	Sequence 44768, A
45	6	30.0	328	5	US-09-979-932A-606	Sequence 606, App

ALIGNMENTS

RESULT 1
US-10-651-165-42
; Sequence 42, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-42

Query Match 100.0%; Score 20; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDCFRKHPDATYRCGSGPW 20
Db 1 TDCFRKHPDATYRCGSGPW 20
|||||

RESULT 2
US-10-651-165-250
; Sequence 250, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19

Query Match 100.0%; Score 20; DB 1; Length 637;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20
|||||
DB 411 TDCFRKHPDATYRCGSGPW 430

RESULT 7
PCT-US03-19834-2
; Sequence 2, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; TITLE OF INVENTION: METHODS FOR TREATING HYPERLIPIDEMIA
; FILE REFERENCE: IOWA:045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-2

Query Match 100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 6.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20
|||||
DB 583 TDCFRKHPDATYRCGSGPW 602

RESULT 8
US-10-651-165-264
; Sequence 264, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 264
; LENGTH: 70
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-264

Query Match 60.0%; Score 12; DB 6; Length 70;
Best Local Similarity 100.0%; Pred. No. 3.3e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATY 12
|||||

Db 15 TDCFRKHPDATY 26

RESULT 9
US-10-296-734-486
; Sequence 486, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 486
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 40
US-10-296-734-486

Query Match 55.0%; Score 11; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 ATYSRCGSGPW 20
|||||
DB 9 ATYSRCGSGPW 19

RESULT 10
US-10-651-165-252
; Sequence 252, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 252
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-252

Query Match 55.0%; Score 11; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 0.00029;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 ATYSRCGSGPW 20
|||||
DB 22 ATYSRCGSGPW 32

RESULT 11
US-10-296-734-816
; Sequence 816, Application US/10296734
; GENERAL INFORMATION:

; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 816
; LENGTH: 1997
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette C
US-10-296-734-816

Query Match 55.0%; Score 11; DB 6; Length 1997;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 ATYSRCGSGPW 20
|||||
Db 1839 ATYSRCGSGPW 1849

RESULT 12
US-10-296-734-406
; Sequence 406, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la consensus polyprotein
US-10-296-734-406

Query Match 55.0%; Score 11; DB 6; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.003;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 ATYSRCGSGPW 20
|||||
Db 592 ATYSRCGSGPW 602

RESULT 13
US-10-296-734-810
; Sequence 810, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 810

; LENGTH: 5985
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la savine
US-10-296-734-810

Query Match 55.0%; Score 11; DB 6; Length 5985;
Best Local Similarity 100.0%; Pred. No. 0.0045;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 ATYRCGSGPW 20
|||||
Db 5829 ATYRCGSGPW 5839

RESULT 14
US-10-651-165-251
; Sequence 251, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 251
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-251

Query Match 50.0%; Score 10; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 ATYRCGSGPW 19
|||||
Db 22 ATYRCGSGPW 31

RESULT 15
US-10-651-165-263
; Sequence 263, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 263
; LENGTH: 70

; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-263

Query Match 45.0%; Score 9; DB 6; Length 70;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPD 9
|||
Db 15 TDCFRKHPD 23

Search completed: November 21, 2003, 22:12:58
Job time : 8.55 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:46:27 ; Search time 9.25 Seconds
(without alignments)
207.932 Million cell updates/sec

Title: US-09-973-025-86

Perfect score: 20

Sequence: 1 TDCFRKHPDARYSRGSGPW 20

Scoring table:

OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR 76:*
1: Piri:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	3011	1 GNVWC3	genome polyprotein
2	12	60.0	3014	1 JCS620	genome polyprotein
3	12	60.0	3033	1 GNVVJ8	genome polyprotein
4	11	55.0	640	2 JQ1584	genome polyprotein
5	11	55.0	716	2 JQ1366	polyprotein - hepa
6	11	55.0	3011	1 S40770	genome polyprotein
7	10	50.0	3011	1 GNVVCH	genome polyprotein
8	9	45.0	3033	1 JQ1303	genome polyprotein
9	8	40.0	234	2 S32742	genome polyprotein
10	8	40.0	235	2 S32747	genome polyprotein
11	8	40.0	237	2 S32744	genome polyprotein
12	8	40.0	350	2 S35631	genome polyprotein
13	8	40.0	782	2 S19876	genome polyprotein
14	8	40.0	782	2 S18031	genome polyprotein
15	8	40.0	782	2 S18032	genome polyprotein
16	8	40.0	782	2 S19875	genome polyprotein
17	8	40.0	787	2 PN0677	hypothetical prote
18	8	40.0	3010	1 GNVVTC	genome polyprotein
19	8	40.0	3010	1 GNVVJC	genome polyprotein
20	8	40.0	3010	1 A45573	genome polyprotein
21	8	40.0	3010	1 S18030	genome polyprotein
22	8	40.0	3010	1 GNVVTV	genome polyprotein
23	6	30.0	114	1 NRUSU2	ribonuclease U2 (
24	6	30.0	280	2 E96997	D-amino acid amino
25	6	30.0	288	2 C72407	hypothetical prote
26	6	30.0	310	2 B55053	endothelial monocy
27	6	30.0	389	2 T03691	calreticulin - com
28	6	30.0	412	2 T05703	calreticulin - bar
29	6	30.0	415	2 T10172	calreticulin - cas

ALIGNMENTS

RESULT 1

GNVWC3

genome polyprotein - hepatitis C virus (strain HCV-1)

N:Contains: capsid protein C; envelope protein M; hepatitisvirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C>Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 19-Jan-2001

C:Accession: A39166; PQ0403; PQ0404

R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Co

Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991

A:Title: Genetic organization and diversity of the hepatitis C virus.

A:Reference number: A39166; MUID:91172826; PMID:1848704

A:Accession: A39166

A:Molecule type: mRNA

A:Residues: 1-3011 <CHO>

A:Cross-references: GB:M2321; NID:g329873; PIDN:AAA45676.1; PID:g329874

R.Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.

J. Gen. Virol. 73, 1131-1141, 1992

A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to

A:Reference number: PQ0393; MUID:92268871; PMID:1316939

A:Accession: PQ0403

A:Molecule type: Genomic RNA

A:Residues: 1577-1633 <CHA>

A:Cross-references: DDBJ:D10128

A:Experimental source: isolates E-b16

A:Accession: PQ0404

A>Status: Preliminary

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CH2>

A:Experimental source: isolates E-b17

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura

F:1-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <ME>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitisvirin #status predicted <NS3>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 100.0%; Score 20; DB 1; Length 3011;

Best Local Similarity 100.0%; Pred. No. 8.5e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDARYSRGSGPW 20

|||||

Db 583 TDCFRKHPDATYRCGSGPW 602

RESULT 2

JC5620
genome polyprotein - hepatitis C virus (isolate EUH1480)
N/Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C/Species: hepatitis C virus
C/Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C/Accession: JC5620
R/Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
Biochem. Biophys. Res. Commun. 236, 44-49, 1997
A/Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant
A/Reference number: JC5620; MUID:97366593; PMID:9223423
A/Accession: JC5620
A/Molecule type: mRNA
A/Residues: 1-3014 <CHA>
A/Cross-references: GB:Y13184
A/Experimental source: genotype 5a, which predominates in South Africa
A/Note: The translation of the nucleotide sequence is not complete in this paper
C/Superfamily: hepatitis C virus genome polyprotein
C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F/2-115/Product: capsid protein C #status predicted <CPC>
F/116-191/Product: major envelope protein M #status predicted <EPW>
F/192-389/Product: major envelope protein E #status predicted <MEB>
F/384-408/Region: hypervariable #status predicted
F/390-730/Product: nonstructural protein NS1 #status predicted <NS1>
F/731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
F/1008-1616/Product: hepatitis C virus genome polyprotein NS2 #status predicted <NS2>
F/1231-1238/Region: hepatitis C virus genome polyprotein NS2 #status predicted <NS2>
F/1313-1318/Region: nucleotide-binding motif A (P-loop)
F/1317-1320/Region: DEXH motif
F/1617-1863/Product: nonstructural protein NS4a #status predicted <N4A>
F/1864-2014/Product: nonstructural protein NS4b #status predicted <N4B>
F/2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F/2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 60.0%; Score 12; DB 1; Length 3014;

Best Local Similarity 100.0%; Pred. No. 4.5e-05; Mismatches 0; Indels 0; Gaps 0;

Matches 12; Conservative 0;

QY 1 TDCFRKHPDATY 12

|||||

Db 584 TDCFRKHPDATY 595

RESULT 3

GNWVJ8
genome polyprotein - hepatitis C virus (strain HC-J8)
N/Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C/Species: hepatitis C virus
C/Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C/Accession: A0250; PQ0397; PQ0559
R/Okamoto, H.; Kura, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.;
Virology 188, 331-341, 1992
A/Title: Full-length sequence of a hepatitis C virus genome having poor homology to rep
A/Reference number: A0250; MUID:92230232; PMID:1314459
A/Accession: A0250

A/Molecule type: genomic RNA

A/Residues: 1-3033 <OKA>

A/Cross-references: GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1; PID:g221609
R/Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.I.
J. Gen. Virol. 73, 1131-1141, 1992

A/Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e

A/Reference number: PQ0393; MUID:92268871; PMID:1316939

A/Accession: PQ0397

A/Molecule type: genomic RNA

A/Residues: 2678-2754 <CHA>

A/Cross-references: DDBJ:D10134

A/Experimental source: isolate E-b12

R/Kato, N.; Ohtsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno

Biochem. Biophys. Res. Commun. 181, 279-285, 1991
A/Title: Distribution of plural HCV types in Japan.
A/Reference number: PQ0554; MUID:92068204; PMID:1720309
A/Accession: PQ0559
A/Molecule type: mRNA

A/Residues: 2678-2729 <KAT>

A/Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524
C/Superfamily: hepatitis C virus genome polyprotein
C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F/1-115/Product: capsid protein C #status predicted <CPC>
F/116-191/Product: major envelope protein M #status predicted <EPW>
F/192-389/Product: major envelope protein E #status predicted <MEB>

F/390-733/Product: nonstructural protein NS1 #status predicted <NS1>

F/734-1010/Product: nonstructural protein NS2 #status predicted <NS2>

F/1011-1619/Product: hepatitis C virus genome polyprotein NS2 #status predicted <NS2>

F/1234-1241/Region: nucleotide-binding motif A (P-loop)

F/1316-1321/Region: nucleotide-binding motif B

F/1320-1323/Region: DEXH motif

F/1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>

F/1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>

F/2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>

F/196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,232

Query Match 60.0%; Score 12; DB 1; Length 3033;

Best Local Similarity 100.0%; Pred. No. 4.6e-05; Mismatches 0; Indels 0; Gaps 0;

Matches 12; Conservative 0;

QY 1 TDCFRKHPDATY 12

|||||

Db 587 TDCFRKHPDATY 598

RESULT 4

JQ1584
genome polyprotein - hepatitis C virus (strain U.K.) (fragment)
N/Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural pro
C/Species: hepatitis C virus
C/Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
C/Accession: JQ1584
R/Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.
J. Gen. Virol. 73, 1521-1525, 1992
A/Title: Cloning and sequencing of the structural region and expression of putative cor
A/Reference number: JQ1584; MUID:92300349; PMID:1318944
A/Accession: JQ1584
A/Molecule type: genomic RNA
A/Residues: 1-640 <KUM>
A/Cross-references: GB:X84079; NID:g643119; PIDN:CAA58888.1; PID:g643120
C/Superfamily: hepatitis C virus genome polyprotein
C/Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; polypr
F/1-191/Product: core protein C #status predicted <CPC>
F/192-389/Product: envelope protein E1 #status predicted <E1>
F/390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predicted
F/196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn) (cov.

Query Match 55.0%; Score 11; DB 2; Length 640;

Best Local Similarity 100.0%; Pred. No. 0.00016; Mismatches 0; Indels 0; Gaps 0;

Matches 11; Conservative 0;

QY 10 ATYSRCGSGPW 20

|||||

Db 592 ATYSRCGSGPW 602

RESULT 5

JQ1366
polyprotein - hepatitis C virus (French isolate) (fragments)
C/Species: hepatitis C virus
C/Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C/Accession: JQ1366
R/Kremsdorff, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.
J. Gen. Virol. 72, 2557-2561, 1991
A/Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implicatio
A/Reference number: JQ1366; MUID:92013977; PMID:1655961

R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A;Accession: Genomic structure of the human prototype strain H of hepatitis C virus: C
A;Reference number: A36814
A;Accession: A36814
A;Molecule type: genomic RNA
A;Residues: 1-3011 <NC>
A;Cross-references: GB:D67463; NID:G329737; PIDN:AAA4534.1; PID:G329738
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari
A;Reference number: A41546; MUID:92052256; PMID:1658800
A;Contents: annotation
A;Note: neither amino acid nor nucleotide sequence is given
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepatitis C virus #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1616-1862/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23
Query Match 50.0%; Score 10; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 0.0069;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 10 ATYSRCGSGP 19
| | | | | | | | | |
Db 592 ATYSRCGSGP 601
RESULT 8
QJ1303
Genome polyprotein - hepatitis C virus (isolate HC-J6)
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (isolate HC-J6) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
C;Accession: QJ1303
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Izuka, H.; Machida, A.; Miyakawa, Y.
J. Gen. Virol. 72, 2697-2704, 1991
A;Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a hum
A;Reference number: QJ1303; MUID:92044440; PMID:1658196
A;Accession: QJ1303
A;Molecule type: genomic RNA
A;Residues: 1-3033 <NA>
A;Cross-references: GB:D00944; NID:G221650; PIDN:BAA00792.1; PID:G221651
A;Experimental source: isolate HC-J6 from a Japanese individual
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; p-loop; polyprotein; serine proteinase; transmem
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F;1011-1619/Product: hepatitis C virus #status predicted <NS3>
F;1316-1321/Region: nucleotide-binding motif B
F;1320-1323/Region: DEXH motif
F;1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
F;1620-1866/Product: nonstructural protein NS4b #status predicted <N4B>
F;1867-2017/Product: nonstructural protein NS5 #status predicted <NS5>
F;2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28
Query Match 45.0%; Score 9; DB 1; Length 3033;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPD 9
|||||

Db 587 TDCFRKHPD 595

RESULT 9

S32742
genome polyprotein - hepatitis C virus (isolate CR-1) (fragment)
N;Contains: envelope protein E2
C;Species: hepatitis C virus
A;Variety: isolate CR-1
C;Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C;Accession: S32742
R;Roggendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
submitted to the EMBL Data Library, February 1993
A;Description: Variability of the envelope regions of HCV in European isolates and its
A;Reference number: S32741
A;Accession: S32742
A;Molecule type: genomic RNA
A;Residues: 1-234 <ROG>
A;Cross-references: EMBL:X72979; NID:g296102; PIDN:CAA51484.1; PID:g296103
A;Experimental source: isolate CR-1
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: envelope protein; polyprotein
F;1-234/Product: envelope protein E2 #status predicted <MAT>

Query Match 40.0%; Score 8; DB 2; Length 234;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHP 8
|||||

Db 214 TDCFRKHP 221

RESULT 10

S32747
genome polyprotein - hepatitis C virus (isolate HU-1) (fragment)
N;Contains: envelope protein E2
C;Species: hepatitis C virus
A;Variety: isolate HU-1
C;Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C;Accession: S32747
R;Roggendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
submitted to the EMBL Data Library, February 1993
A;Description: Variability of the envelope regions of HCV in European isolates and its
A;Reference number: S32741
A;Accession: S32747
A;Molecule type: genomic RNA
A;Residues: 1-235 <ROG>
A;Cross-references: EMBL:X72977; NID:g296112; PIDN:CAA51482.1; PID:g296113
A;Experimental source: isolate HU-1
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: envelope protein; polyprotein
F;1-235/Product: envelope protein E2 #status predicted <MAT>

Query Match 40.0%; Score 8; DB 2; Length 235;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHP 8
|||||

Db 215 TDCFRKHP 222

RESULT 11

S32744
genome polyprotein - hepatitis C virus (isolate EG-1) (fragment)
N;Contains: envelope protein E2
C;Species: hepatitis C virus
A;Variety: isolate EG-1

C;Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C;Accession: S32744
R;Roggendorf, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
submitted to the EMBL Data Library, February 1993
A;Description: Variability of the envelope regions of HCV in European isolates and its
A;Reference number: S32741
A;Accession: S32744
A;Molecule type: genomic RNA
A;Residues: 1-237 <ROG>
A;Cross-references: EMBL:X72981
A;Experimental source: isolate EG-1
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: envelope protein; polyprotein
F;1-237/Product: envelope protein E2 #status predicted <MAT>

Query Match 40.0%; Score 8; DB 2; Length 237;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHP 8
|||||

Db 217 TDCFRKHP 224

RESULT 12

S35631
genome polyprotein - hepatitis C virus (fragment)
N;Contains: nonstructural protein 1
C;Species: hepatitis C virus
C;Date: 19-Mar-1997 #sequence_revision 30-Jan-1998 #text_change 17-Nov-2000
C;Accession: S35631
R;Sarashina, T.; Sakurai, T.; Watanabe, Y.; Kashima, K.; Suzuki, T.; Chiba, J.; Kita, Y.
Nucleic Acids Res. 21, 1037, 1993
A;Title: Nucleotide sequence of the hepatitis C virus genome from a patient negative for
A;Reference number: S35629; MUID:93197128; PMID:8383835
A;Accession: S35631
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-350 <SAR>
A;Cross-references: EMBL:D13970
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1992
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: polyprotein
F;1-350/Product: nonstructural protein 1 #status predicted <MAT>

Query Match 40.0%; Score 8; DB 2; Length 350;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHP 8
|||||

Db 204 TDCFRKHP 211

RESULT 13

S19876
genome polyprotein - hepatitis C virus (isolate JK5) (fragment)
N;Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C;Species: hepatitis C virus
A;Variety: isolate JK5
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C;Accession: S19876
R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A;Description: Sequence analysis of putative structural regions of Hepatitis C Virus is
A;Reference number: S18029
A;Accession: S19876
A;Molecule type: genomic RNA
A;Residues: 1-782 <HON>
A;Cross-references: EMBL:X61595; NID:g59486; PIDN:CAA43792.1; PID:g59487
A;Experimental source: isolate JK5
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural

F;1-191/Product: core protein #status predicted <MAT1>
F;192-383/Product: envelope protein 1 #status predicted <MAT2>
F;384-733/Product: NS1/E2 protein #status predicted <MAT3>
F;734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 40.0%; Score 8; DB 2; Length 782;
Best Local Similarity 100.0%; Pred. No. 0.36; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDCFRKHP 8
|||
Db 583 TDCFRKHP 590

RESULT 14

S18031
genome polypeptide - hepatitis C virus (isolate JK2) (fragment)
N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C:Species: hepatitis C virus
A:Variety: isolate JK2
C:Accession: S18031
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
R:Honda, M.; Kaneko, S.; Maeshashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
A:Reference number: S18029
A:Accession: S18031
A:Molecule type: genomic RNA
A:Residues: 1-782 <HON>
A:Cross-references: EMBL:X61593
C:Experimental source: isolate JK2
C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F;1-191/Product: core protein #status predicted <MAT1>
F;192-383/Product: envelope protein 1 #status predicted <MAT2>
F;384-733/Product: NS1/E2 protein #status predicted <MAT3>
F;734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 40.0%; Score 8; DB 2; Length 782;
Best Local Similarity 100.0%; Pred. No. 0.36; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDCFRKHP 8
|||
Db 583 TDCFRKHP 590

RESULT 15

S18032
genome polypeptide - hepatitis C virus (isolate JK4) (fragment)
N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C:Species: hepatitis C virus
A:Variety: isolate JK4
C:Accession: S18032
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
R:Honda, M.; Kaneko, S.; Maeshashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
A:Reference number: S18029
A:Accession: S18032
A:Molecule type: genomic RNA
A:Residues: 1-782 <HON>
A:Cross-references: EMBL:X61594
C:Experimental source: isolate JK4
C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F;1-191/Product: core protein #status predicted <MAT1>
F;192-383/Product: envelope protein 1 #status predicted <MAT2>
F;384-733/Product: NS1/E2 protein #status predicted <MAT3>
F;734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 40.0%; Score 8; DB 2; Length 782;
Best Local Similarity 100.0%; Pred. No. 0.36; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TDCFRKHP 8
|||
Db 583 TDCFRKHP 590

Search completed: November 21, 2003, 21:11:36
Job time : 10.25 secs

THIS PAGE BLANK (HSPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:52 ; Search time 4.85 Seconds
(without alignments)
193.925 Million cell updates/sec

Title: US-09-973-025-86

Perfect score: 20

Sequence: 1 TDCFRKHDPATYRCGSGFW 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query	ID	Description
1	20	100.0	3011	1	POLG_HCV1
2	12	60.0	3033	1	POLG_HCVJ8
3	10	50.0	3011	1	POLG_HCVH
4	9	45.0	737	1	POLG_HCVJ7
5	9	45.0	3033	1	POLG_HCVJ6
6	8	40.0	3010	1	POLG_HCVBK
7	8	40.0	3010	1	POLG_HCVJA
8	8	40.0	3010	1	POLG_HCVJT
9	8	40.0	3010	1	POLG_HCVTW
10	7	35.0	424	1	CRTC_ORYSA
11	7	35.0	737	1	POLG_HCVJ5
12	6	30.0	114	1	RNU2_USTSP
13	6	30.0	312	1	MCAL_HUMAN
14	6	30.0	367	1	PROB_PASMU
15	6	30.0	374	1	PROB_RALSO
16	6	30.0	415	1	CRTC_RICCO
17	6	30.0	416	1	CRTC_BETVU
18	6	30.0	416	1	CRTC_NICPL
19	6	30.0	420	1	CRTC_WAIZE
20	6	30.0	421	1	CRTC_PRUAR
21	6	30.0	424	1	CRTC_ARATH
22	6	30.0	425	1	CRTC_ARATH
23	6	30.0	601	1	FL3C_MOUSE
24	6	30.0	676	1	CMC2_MOUSE
25	6	30.0	758	1	PARC_RHIME
26	6	30.0	959	1	SYL_AERPE
27	5	25.0	34	1	AMP2_CHICK
28	5	25.0	38	1	HMG2_BOVIN
29	5	25.0	39	1	AMP1_CHICK
30	5	25.0	39	1	GLL2_CHICK
31	5	25.0	65	1	GLL1_CHICK
32	5	25.0	78	1	RL31_RICCN
33	5	25.0	78	1	RL31_RICPR

Q15649 homo sapien
Q27371 methanobact
Q9ugv6 homo sapien
P10103 bos taurus
P09429 homo sapien
P07155 mus musculus
P12682 sus scrofa
P28941 equine herp
Q03077 osmunda cin
Q926a7 listeria in
Q928d5 listeria in
P16650 potato viru

ALIGNMENTS

RESULT 1

POLG_HCV1
ID POLG_HCV1 STANDARD; PRT; 3011 AA.
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate 1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Colt D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;
RT "Genetic organization and diversity of the hepatitis C virus."; Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA}(N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; M62321; AAA45676.1; -.
CC PIR; A39166; GNMVVC3.
CC PDB; 1A1V; 16-FEB-99.
CC PDB; 1HEI; 25-NOV-98.
CC MEROPS; S29.001; -.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002522; HCV_capsid.

DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002186; HCV_RdRP.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRP; 1.
 DR Pfam; PF0186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDc7; 1.
 KW Polyprotein; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115
 FT CHAIN 116 191 MATRIX PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH_BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FC5AE9 CRC64;
 Query Match 100.0%; Score 20; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 1.9e-14;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYSRGSGPW 20

Db 583 TDCFRKHPDATYSRGSGPW 602

RESULT 2

POLG_HCVJ8 STANDARD; PRT; 3033 AA.

AC P26661; 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J8) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 CC NCBI_TaxID=11115;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kural K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
 RA Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes."; RT
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC [RNA] (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D10988; BAA01761.1; -.
 CC PIR; A40250; GNMVJ8.
 CC HSP; P27958; 1HEI.
 CC MEROPS; S29.001; -.
 CC DR InterPro; IPR001410; DEAD.
 CC InterPro; IPR002522; HCV_capsid.
 CC InterPro; IPR002521; HCV_core.
 CC InterPro; IPR002519; HCV_env.
 CC InterPro; IPR002531; HCV_NS1.
 CC InterPro; IPR002518; HCV_NS2.
 CC InterPro; IPR004109; HCV_NS3.
 CC InterPro; IPR000745; HCV_NS4a.
 CC InterPro; IPR001490; HCV_NS4b.
 CC InterPro; IPR002868; HCV_NS5a.
 CC InterPro; IPR002166; HCV_RdRP.
 CC InterPro; IPR007095; RNA_pol_DS_PS.
 CC InterPro; IPR007094; RNA_pol_PSVir.
 CC Pfam; PF01543; HCV_capsid; 1.
 CC Pfam; PF01542; HCV_core; 1.


```

DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00998; Viral RGRP; 1.
DR Pfam; PF018062; HCV NS1; 1.
DR Pfam; PF00487; DEXdc; 1.
DR SMART; SM00487; DEXdc; 1.
KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolyase; Serine protease.
KW TRANSMEMBRANE; NONSTRUCTURAL PROTEIN C BY THE
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT CHAIN 3033 369
FT TRANSMEM 347
FT ACT SITE 1087
FT ACT SITE 1111
FT ACT SITE 1169
FT NP_BIND 1234
FT SITE 1320
FT CARBOHYD 196
FT CARBOHYD 209
FT CARBOHYD 233
FT CARBOHYD 299
FT CARBOHYD 305
FT CARBOHYD 417
FT CARBOHYD 423
FT CARBOHYD 430
FT CARBOHYD 448
FT CARBOHYD 477
FT CARBOHYD 534
FT CARBOHYD 542
FT CARBOHYD 558
FT CARBOHYD 578
FT CARBOHYD 627
FT CARBOHYD 649
FT CARBOHYD 1091
FT CARBOHYD 2038
FT CARBOHYD 2359
FT CARBOHYD 2811
SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;

Query Match 60.0%; Score 12; DB 1; Length 3033;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDCFRKHPDQY 12
DB 587 TDCFRKHPDQY 598

RESULT 3
ID POLG HCWH STANDARD; PRT; 3011 AA.
AC P27958.
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein

NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
Hepatitis C virus (isolate H) (HCV).
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus
NCBI_TaxID=11108;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658800;
RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C
RT virus: comparison with American and Japanese isolates.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN (2)
RX X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX MEDLINE=97331322; PubMed=9187654;
RA Yao N., Hesson I., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain.";
RL Nat. Struct. Biol. 4:463-467(1997).
RN (3)
RX X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE=98154321; PubMed=9493270;
RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding.";
RL Structure 6:99-100(1998).
CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M67463; AAA45534.1; -.
PIR; A36814; GNVVCH.
PDB; 1HEI; 25-NOV-98.
PDB; 1A1V; 16-FEB-99.
PDB; 1A1R; 17-JUN-98.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR TRANSFAC; T04155; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.

```

DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002888; HCV NS5a.
 DR InterPro: IPR002166; HCV RdRP.
 DR InterPro: IPR001650; Helicase C.
 DR InterPro: IPR007095; RNA pol Ds ps.
 DR InterPro: IPR007094; RNA pol PSvir.
 DR Pfam: PF01543; HCV capsid; 1.
 DR Pfam: PF01542; HCV core; 1.
 DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR Pfam: PF01538; HCV NS2; 1.
 DR Pfam: PF02307; HCV NS3; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF01506; HCV NS5a; 1.
 DR Pfam: PF00271; helicase C; 1.
 DR Pfam: PF00998; Viral RdRP; 1.
 DR ProDom: PD186062; HCV NS1; 1.
 DR SMART: SM00487; DEXdc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 191
 FT CHAIN 192 383 ENVELOPE GLYCOPROTEIN E1.
 FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E2.
 FT CHAIN 747 809 PROTEIN P7.
 FT CHAIN 810 1026 NONSTRUCTURAL PROTEIN NS2.
 FT CHAIN 1027 1657 PROTEASE/HELICASE NS3.
 FT CHAIN 1658 1711 NONSTRUCTURAL PROTEIN NS4A.
 FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4B.
 FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS5A.
 FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5B.
 FT CHAIN 3012 369 POTENTIAL.
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1224 1226
 FT TURN 1232 1233
 FT TURN 1236 1238
 FT TURN 1239 1246
 FT TURN 1247 1248
 FT STRAND 1251 1255
 FT TURN 1258 1271
 FT TURN 1272 1272
 FT TURN 1277 1280
 FT TURN 1281 1282
 FT STRAND 1283 1285
 FT STRAND 1291 1295
 FT TURN 1296 1301
 FT TURN 1302 1303
 FT TURN 1312 1316
 FT TURN 1317 1319

FT HELIX 1323 1335
 FT TURN 1336 1340
 FT STRAND 1343 1347
 FT TURN 1352 1353
 FT TURN 1360 1361
 FT STRAND 1362 1366
 FT STRAND 1368 1368
 FT STRAND 1373 1375
 FT TURN 1376 1377
 FT STRAND 1378 1380
 FT HELIX 1382 1385
 FT STRAND 1389 1393
 FT HELIX 1397 1409
 FT TURN 1410 1411
 FT STRAND 1414 1417
 FT TURN 1419 1420
 FT STRAND 1432 1436
 FT TURN 1438 1439
 FT STRAND 1450 1453
 FT STRAND 1456 1463
 FT STRAND 1471 1478
 FT STRAND 1480 1480
 FT HELIX 1481 1488
 FT TURN 1489 1490
 FT STRAND 1497 1501
 FT STRAND 1507 1507
 FT STRAND 1511 1511
 FT HELIX 1514 1527
 FT STRAND 1532 1544
 FT STRAND 1550 1550
 FT HELIX 1555 1564
 FT HELIX 1570 1578
 FT TURN 1579 1580
 FT TURN 1598 1598
 FT HELIX 1584 1597
 FT HELIX 1606 1611
 FT TURN 1614 1618
 FT STRAND 1622 1623
 FT STRAND 1627 1627
 FT STRAND 1635 1636
 FT HELIX 1640 1652
 SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;
 Query Match 50.0%; Score 10; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 0.0026;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 10 ATYRCGSGP 19
 |||||
 Db 592 ATYRCGSGP 601
 |||||
 RESULT 4
 POLG HCVJ7 STANDARD; PRT; 737 AA.
 ID POLG HCVJ7
 AC P27961;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
 protein (Envelope protein M); Major envelope protein E; Nonstructural
 proteins NS1 and NS2] (Fragment).
 DE Hepatitis C virus (isolate HC-J7) (HCV).
 OS Hepatitis C virus; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NCBI_TaxID=11114;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 homology to reported isolates: comparative study of four distinct

RT genotypes";

RL Virology 188:331-341(1992).

CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE

CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.

CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: OF

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND MRNA.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; D10077; BAA00971.1; -

DR InterPro; IPR002522; HCV capsid.

DR InterPro; IPR002521; HCV core.

DR InterPro; IPR002519; HCV env.

DR InterPro; IPR002531; HCV NS1.

DR Pfam; PF01543; HCV capsid; 1.

DR Pfam; PF01542; HCV core; 1.

DR Pfam; PF01539; HCV env; 1.

DR Pfam; PF01560; HCV NS1; 1.

DR Pfam; PF01538; HCV NS2; 1.

DR Pfam; PF02907; HCV NS3; 1.

DR Pfam; PF01006; HCV NS4a; 1.

DR Pfam; PF01001; HCV NS4b; 1.

DR Pfam; PF01506; HCV NS5a; 1.

DR Pfam; PF00271; helicase C; 1.

DR Pfam; PF00998; Viral RdRP; 1.

DR ProDom; PD186062; HCV NS1; 1.

KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;

KW Transmembrane; Nonstructural protein.

FT INIT_MET 1 1

FT CHAIN 1 115

FT CHAIN 116 191

FT CHAIN 192 383

FT CHAIN 384 733

FT CHAIN 734 >737

FT TRANSMEM 347 369

FT CARBOHYD 136 196

FT CARBOHYD 209 209

FT CARBOHYD 233 233

FT CARBOHYD 299 299

FT CARBOHYD 305 305

FT CARBOHYD 417 417

FT CARBOHYD 423 423

FT CARBOHYD 430 430

FT CARBOHYD 448 448

FT CARBOHYD 477 477

FT CARBOHYD 534 534

FT CARBOHYD 542 542

FT CARBOHYD 558 558

FT CARBOHYD 578 578

FT CARBOHYD 627 627

FT CARBOHYD 649 649

FT NON_TER 737 737

SQ SEQUENCE 737 AA; 81691 MW; 67DPAE11854122F2 CRC64;

Query Match 45.0%; Score 9; DB 1; Length 737;

Best Local Similarity 100.0%; Pred. No. 0.0097;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPD 9

|||||

Db 587 TDCFRKHPD 595

RESULT 5

POLG HCVJ6

ID POLG HCVJ6 STANDARD; PRT; 3033 AA.

AC P26660:

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);

DE Envelope glycoprotein E1 (GP32) (P35); Envelope glycoprotein E2

DE (GP68) (P70) (NS1); Protein P7; Nonstructural protein NS2 (P21)

DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)

DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein

DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein

DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

OS Hepatitis C virus (isolate HC-J6) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OC NCBI_TaxID=11113;

OX [1]

RN SEQUENCE FROM N.A.

RP MEDLINE=92044440; PubMed=1658196;

RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,

RA Machida A., Miyakawa Y., Mayumi M.;

RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated

RT from a human carrier: comparison with reported isolates for conserved

RT and divergent regions.";

RL J. Gen. Virol. 72:2697-2704(1991).

CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE

CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.

CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral

CC precursor polyprotein, commonly with Asp or Glu in the P6

CC position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate +

CC {RNA}(N).

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND MRNA.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; D00944; BAA00792.1; -

DR PIR; JQ1303; JQ1303.

DR HSSP; P27958; 1HEI.

DR MEROPS; S29.001; -

DR MEROPS; U39.001; -

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV capsid.

DR InterPro; IPR002521; HCV core.

DR InterPro; IPR002519; HCV env.

DR InterPro; IPR002531; HCV NS1.

DR InterPro; IPR002518; HCV NS2.

DR InterPro; IPR004109; HCV NS3.

DR InterPro; IPR000745; HCV NS4a.

DR InterPro; IPR001490; HCV NS4b.

DR InterPro; IPR002868; HCV NS5a.

DR InterPro; IPR002166; HCV RdRP.

DR InterPro; IPR001850; Helicase C.

DR InterPro; IPR007095; RNA pol_D5_P5.

DR InterPro; IPR007094; RNA pol_P5vir.

DR Pfam; PF01543; HCV capsid; 1.

DR Pfam; PF01542; HCV core; 1.

DR Pfam; PF01539; HCV env; 1.

DR Pfam; PF01560; HCV NS1; 1.

DR Pfam; PF01538; HCV NS2; 1.

DR Pfam; PF02907; HCV NS3; 1.

DR Pfam; PF01006; HCV NS4a; 1.

DR Pfam; PF01001; HCV NS4b; 1.

DR Pfam; PF01506; HCV NS5a; 1.

DR Pfam; PF00271; helicase C; 1.

DR Pfam; PF00998; Viral RdRP; 1.

DR ProDom; PD186062; HCV NS1; 1.

DR SMART; SM00487; DEXDC; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2811 2811
 SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;

Query Match 45.0%; Score 9; DB 1; Length 3033;
 Best Local Similarity 100.0%; Pred. No. 0.034; 0; Indels 0; Gaps 0;
 Matches 9; Conservative 0; Mismatches 0; Gaps 0;
 QY 1 TDCFRKHPD 9
 |||||
 Db 587 TDCFRKHPD 595

RESULT 6
 POLG_HCVBK STANDARD; PRT; 3010 AA.
 AC P26663;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polypeptide [Contains: Capsid protein c (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate BK) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11105;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91140698; PubMed=1847440;
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 Onishi E., Andoh T., Yoshida I., Okayama H.;

RT "Structure and organization of the hepatitis C virus genome isolated
 from human carriers.";
 J. Virol. 65:1105-1113(1991).
 RN [2]
 RP SEQUENCE OF 1487-1500.
 RX MEDLINE=96235224; PubMed=8647104;
 RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetetsky L.;
 RT "Non-structural protein 3 of hepatitis C virus inhibits
 phosphorylation mediated by cAMP-dependent protein kinase.";
 Eur. J. Biochem. 237:611-618(1996).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RX MEDLINE=97015088; PubMed=9861916;
 RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
 RA Moosaw E.W., Adachi T., Hostomsky Z.;
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 trypsin-like fold and a structural zinc binding site.";
 Cell 87:331-342(1996).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RX MEDLINE=98227846; PubMed=9568891;
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
 virus: a 2.2-A resolution structure in a hexagonal crystal form.";
 Protein Sci. 7:837-847(1998).
 RN [5]
 RP FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polypeptide, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M58335; AAA72945.1; -.
 CC PIR; A38465; GNVVTC
 CC PDB; 1AIQ; 25-MAR-98.
 CC PDB; 1JXP; 14-JAN-98.
 CC PDB; 1NS3; 08-APR-98.
 CC PDB; 1C2P; 15-NOV-00.
 CC PDB; 1CSJ; 08-NOV-99.
 CC PDB; 1CX5; 09-APR-02.
 CC PDB; 1XG6; 10-APR-02.
 CC PDB; 1QUV; 26-JUN-00.
 CC PDB; 8OHM; 20-APR-99.
 CC MEROPS; S29.001; -.
 CC InterPro; IPR001410; DEAD.
 CC InterPro; IPR002522; HCV_capsid.
 CC InterPro; IPR002521; HCV_core.
 CC InterPro; IPR002519; HCV_env.
 CC InterPro; IPR002531; HCV_NS1.
 CC InterPro; IPR002518; HCV_NS2.
 CC InterPro; IPR004109; HCV_NS3.
 CC InterPro; IPR000745; HCV_NS4a.
 CC InterPro; IPR001490; HCV_NS4b.
 CC InterPro; IPR002868; HCV_NS5a.
 CC InterPro; IPR002166; HCV_RGRP.
 CC InterPro; IPR007095; RNA_pol_DS_PS.

FT	STRAND	1139	1144
FT	STRAND	1149	1157
FT	HELIX	1158	1161
FT	TURN	1162	1163
FT	TURN	1165	1166
FT	TURN	1168	1171
FT	TURN	1172	1174
FT	STRAND	1175	1186
FT	TURN	1187	1188
FT	STRAND	1189	1197
FT	TURN	1198	1202
FT	HELIX	1203	1204
FT	STRAND	1680	1688
SQ	SEQUENCE	3010 AA; 3	3010 AA; 3

		Query Match	
		Best Local Similarity	
		Matches	
		8; Conservati	

Qy	1 TDCPRKHP	8
Db	583 TDCPRKHP	590

RESULT 7		
ID	POLG HCVJA	STANDARD
AC	P26662;	
DT	01-AUG-1992 (Rel. 23)	
DT	01-AUG-1992 (Rel. 23)	
DE	28-FEB-2003 (Rel. 41)	
DE	Genome polyprotein (Co	
DE	Envelope glycoprotein (GP68) (GP70) (NS1); P	
DE	(BC 3.4.22.-); Proteas	
DE	(BC 3.4.21.98); Nonstru	
DE	N54B (P27); Nonstructu	
DE	N5B (P66) (P70) (RNA-	
OS	Hepatitis C virus (iso	
OC	Viruses; ssRNA positiv	
OC	Hepadnavirus.	
OX	NCBI_Taxid=11116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=91089550; PubM	
RA	Kato N., Hijikata M., I	
RA	Ogkoshi S., Shimotohno	
RA	Sugimura T., Shimotohno	
RT	"Molecular cloning of	
RT	Japanese patients with	
RL	Proc. Natl. Acad. Sci.	
RL	[2]	
RP	DISCUSSION OF SEQUENCE	
RP	MEDLINE=91192160; PubMM	
RA	Kato N., Hijikata M., I	
RA	Ogkoshi S., Shimotohno	
RC	"Molecular structure o	
CC	FEBS Lett. 280:325-328	
CC	-I- FUNCTION: THE SWAL	
CC	HYDROPHOBIC, SUGGE	
CC	NS3 AND NS5 MAY PL	
CC	-I- CATALYTIC ACTIVITY	
CC	precursor polyp	
CC	position. Cys or T	
CC	-I- CATALYTIC ACTIVITY	
CC	{RNA} [N].	
CC	-I- SUBUNIT: THE VIRIO	
CC	LIPIDOPROTEIN ENVELO	
CC	PROTEIN M AND GLYC	
CC	PROTEIN C AND WNA	
CC	-I- SIMILARITY: THE PR	
CC	-----	
CC	This SWISS-PROT entry	
CC	between the Swiss Ins	

DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV NS5b.
 DR InterPro; IPR007095; RNA pol PS.
 DR InterPro; IPR007094; RNA pol PSvir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF02271; Helicase C; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1093
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 224 224
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 522 522
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2529 2529
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;

Query Match 40.0%; Score 8; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.43;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHP 8
 |||||
 Db 583 TDCFRKHP 590

RESULT 9
 POLG_HCVTW STANDARD; PRT; 3010 AA.
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus
 OX NCBI_TaxID=31645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230206; PubMed=1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 RT "The Taiwanese hepatitis C virus genome: sequence determination and
 RT mapping the 5' termini of viral genomic and antigenomic RNA.";
 RL Virology 188:102-113(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA)(N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M84754; -; NOT ANNOTATED_CDS.
 CC PIR; A40244; GNVVTW.
 CC PDB; 1N64; 25-FEB-03.
 CC PDB; 1NS3; 08-APR-98.
 CC MEROPS; S29.001; -;
 CC MEROPS; U39.001; -;
 CC InterPro; IPR001410; DEAD.
 CC InterPro; IPR002522; HCV capsid.
 CC InterPro; IPR002521; HCV core.
 CC InterPro; IPR002519; HCV env.
 CC InterPro; IPR002531; HCV NS1.
 CC InterPro; IPR002518; HCV NS2.
 CC InterPro; IPR004109; HCV NS3.
 CC InterPro; IPR000745; HCV NS4a.
 CC InterPro; IPR001490; HCV NS4b.
 CC InterPro; IPR002868; HCV NS5a.
 CC InterPro; IPR002166; HCV RdRP.
 CC InterPro; IPR007095; RNA_pol_DS_PS.
 CC InterPro; IPR007094; RNA_pol_Psvir.
 CC Pfam; PF01543; HCV capsid; 1.
 CC Pfam; PF01542; HCV core; 1.
 CC Pfam; PF01539; HCV env; 1.
 CC Pfam; PF01560; HCV NS1; 1.
 CC Pfam; PF01538; HCV NS2; 1.
 CC Pfam; PF02907; HCV NS3; 1.


```

FT HELIX 16 31
FT TURN 32 33
FT HELIX 36 38
FT STRAND 41 43
FT TURN 60 64
FT TURN 67 68
FT STRAND 74 74
FT TURN 75 76
FT STRAND 77 77
FT STRAND 84 89
FT TURN 90 92
FT STRAND 95 100
FT TURN 102 103
FT TURN 107 108
FT STRAND 111 112
SQ SEQUENCE 114 AA; 12387 MW; 96D32CB2E23AAB98 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 CGSGPW 20
Db 55 CGSGPW 60
|||||

RESULT 13
MCAL HUMAN STANDARD; PRT; 312 AA.
ID MCAL HUMAN STANDARD; PRT; 312 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Multisynthetase complex auxiliary component p43 [Contains:
DE Endothelial-monocyte activating polypeptide II (EMAP-II) (small
DE inducible cytokine subfamily E member 1)].
GN SCYE1 OR EMAP2.
OS Homo. sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95014290; PubMed=7929199;
RA Kao J., Houck K., Fan Y., Haehnel I., Libutti S.K., Kayton M.L.,
RA Grikscheit T., Chabot J., Nowygrod R., Greenberg S., Kuang W.J.,
RA Leung D.W., Hayward J.R., Kisiel W., Heath M., Brett J., Stern D.M.;
RT "Characterization of a novel tumor-derived cytokine. Endothelial-
RT monocyte activating polypeptide II";
RL J. Biol. Chem. 269:25106-25119(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
```

```

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SIMILARITY: Contains 1 tRNA-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U10117; AAA62202.1; -
CC EMBL; BC014051; AAH14051.1; -
CC PDB; 1E7Z; 06-FEB-01.
CC PDB; 1EUJ; 06-SEP-00.
CC PDB; 1FLO; 07-FEB-01.
CC Genew; HGNC:10648; SCYE1.
CC MIM; 603605; -
CC GO; GO:0005615; C:extracellular space; TAS.
CC GO; GO:0005125; F:cytokine activity; TAS.
CC GO; GO:0000049; F:tRNA binding activity; TAS.
CC GO; GO:0006418; P:amino acid activation; TAS.
CC GO; GO:0006935; P:chemotaxis; TAS.
CC GO; GO:0006954; P:inflammatory response; TAS.
CC InterPro; IPR002547; tRNA_bind.
CC Pfam; PF01588; tRNA_bind; 1.
CC PIRSF; PIRSF005381; EMAP11; 1.
CC PROSITE; PS50886; TRBD; 1.
CC Protein biosynthesis; RNA-binding; tRNA-binding; Cytokine;
KW 3D-structure. 1 146
FT PROPEP 147 312
FT CHAIN 146 ENDOTHELIAL-MONOCYTE ACTIVATING
FT POLYPEPTIDE II.
FT DOMAIN 151 252 TRNA-BINDING.
FT CONFLICT 79 A -> P (IN REF. 2).
FT SEQUENCE 312 AA; 34326 MW; 946310A0216F7587 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 RKHPDA 10
Db 166 RKHPDA 171
|||||

RESULT 14
PROB PASMU
ID PROB PASMU STANDARD; PRT; 367 AA.
AC Q9CUJ5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutamate 5-kinase (EC 2.7.2.11) (Gamma-glutamyl kinase) (GK).
GN PROB OR PM1896.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Faustian M.L., Whittam T.S., Kapur V.;
RA "Complete genomic sequence of Pasteurella multocida Pm70.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- FUNCTION: Catalyzes the transfer of a phosphate group to glutamate
CC to form glutamate 5-phosphate which rapidly cyclizes to 5-
CC oxoproline.
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate = ADP + L-glutamate 5-
CC phosphate.
CC -!- PATHWAY: Proline biosynthesis; first step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
```

```
CC -!- SIMILARITY: Belongs to the glutamate 5-kinase family.
CC -!- SIMILARITY: Contains 1 PUA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE006227; AAK03980.1; ALT_INIT.
CC HAMAP; MF_00456; -; 1.
CC InterPro; IPR001048; Aa_kinase.
CC InterPro; IPR001057; Glu_5kinase.
CC InterPro; IPR005715; ProB.
CC InterPro; IPR002478; PUA.
CC Pfam; PF00696; aakina; 1.
CC Pfam; PF01472; PUA; 1.
CC PRINTS; PR00474; GLU5KINASE.
CC SMART; SM00359; PUA; 1.
CC TIGRPFAM; TIGR01027; proB; 1.
CC PROSITE; PS00902; GLUTAMATE_5_KINASE; 1.
CC PROSITE; PS50890; PUA; 1.
CC TRANSFERASE; Kinase; Proline biosynthesis; Complete proteome.
KW DOMAIN 275 353 PUA.
FT SEQUENCE 367 AA; 39894 MW; E62DA632EC78F6E5 CRC64;
SQ
Query Match 30.0%; Score 6; DB 1; Length 367;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 5 RKHPDA 10
Db 180 RKHPDA 185
RESULT 15
PROB_RALSO STANDARD; PRT; 374 AA.
AC Q8XVL1
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutamate 5-kinase (EC 2.7.2.11) (Gamma-glutamyl kinase) (GK).
GN PROB OR RSC2819 OR RS00286.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=216181879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Signier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
CC -!- FUNCTION: Catalyzes the transfer of a phosphate group to glutamate
CC to form glutamate 5-phosphate which rapidly cyclizes to 5-
CC oxoproline.
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate = ADP + L-glutamate 5-
CC phosphate.
CC -!- PATHWAY: Proline biosynthesis; first step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the glutamate 5-kinase family.
CC -!- SIMILARITY: Contains 1 PUA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL646072; CAD16526.1; -.
CC HAMAP; MF_00456; -; 1.
CC InterPro; IPR001048; Aa_kinase.
CC InterPro; IPR001057; Glu_5kinase.
CC InterPro; IPR005715; ProB.
CC InterPro; IPR002478; PUA.
CC Pfam; PF00696; aakina; 1.
CC Pfam; PF01472; PUA; 1.
CC PRINTS; PR00474; GLU5KINASE.
CC SMART; SM00359; PUA; 1.
CC TIGRPFAM; TIGR01027; proB; 1.
CC PROSITE; PS00902; GLUTAMATE_5_KINASE; 1.
CC PROSITE; PS50890; PUA; 1.
CC TRANSFERASE; Kinase; Proline biosynthesis; Complete proteome.
KW DOMAIN 282 360 PUA.
FT SEQUENCE 374 AA; 39366 MW; C2050F5CB6C2C4BC CRC64;
SQ
Query Match 30.0%; Score 6; DB 1; Length 374;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 5 RKHPDA 10
Db 186 RKHPDA 191
Search completed: November 21, 2003, 21:00:00
Job time : 5.85 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:45:57 ; Search time 24.4 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-86
Perfect score: 20
Sequence: 1 TDCFRKHPDATYRCGSGPW 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	137	Q914W1	Q914W1 hepatitis c
2	20	100.0	2436	Q81756	Q81756 hepatitis c
3	20	100.0	3011	Q91FES	Q91FES hepatitis c
4	20	100.0	3011	Q81754	Q81754 hepatitis c
5	16	80.0	3021	Q68870	Q68870 hepatitis c
6	13	65.0	137	Q914S8	Q914S8 hepatitis c
7	12	60.0	350	Q81323	Q81323 hepatitis c
8	12	60.0	350	Q81318	Q81318 hepatitis c
9	12	60.0	350	Q81324	Q81324 hepatitis c
10	12	60.0	350	Q81320	Q81320 hepatitis c
11	12	60.0	350	Q81322	Q81322 hepatitis c
12	12	60.0	350	Q81325	Q81325 hepatitis c
13	12	60.0	350	Q81321	Q81321 hepatitis c
14	12	60.0	350	Q81316	Q81316 hepatitis c
15	12	60.0	350	Q81317	Q81317 hepatitis c
16	12	60.0	350	Q81319	Q81319 hepatitis c

17	12	60.0	3010	12	093077	093077 hepatitis c
18	12	60.0	3010	12	Q8QRL8	Q8QRL8 hepatitis c
19	12	60.0	3014	12	Q91936	Q91936 hepatitis c
20	12	60.0	3014	12	Q39928	Q39928 hepatitis c
21	12	60.0	3019	12	Q68801	Q68801 hepatitis c
22	12	60.0	3033	12	Q9DHD6	Q9DHD6 hepatitis c
23	12	60.0	3033	12	Q991B8	Q991B8 hepatitis c
24	12	60.0	3033	12	Q9Q9A7	Q9Q9A7 hepatitis c
25	12	60.0	3033	12	Q91ZA2	Q91ZA2 hepatitis c
26	12	60.0	3033	12	Q991B3	Q991B3 hepatitis c
27	11	55.0	128	12	Q8JYQ3	Q8JYQ3 hepatitis c
28	11	55.0	128	12	Q8JYQ2	Q8JYQ2 hepatitis c
29	11	55.0	128	12	Q8JYQ4	Q8JYQ4 hepatitis c
30	11	55.0	129	12	Q8JYQ1	Q8JYQ1 hepatitis c
31	11	55.0	129	12	Q8JYQ9	Q8JYQ9 hepatitis c
32	11	55.0	129	12	Q8JYQ0	Q8JYQ0 hepatitis c
33	11	55.0	137	12	Q914U0	Q914U0 hepatitis c
34	11	55.0	137	12	Q914T8	Q914T8 hepatitis c
35	11	55.0	137	12	Q914Q6	Q914Q6 hepatitis c
36	11	55.0	137	12	Q914V3	Q914V3 hepatitis c
37	11	55.0	137	12	Q914U3	Q914U3 hepatitis c
38	11	55.0	137	12	Q914Q5	Q914Q5 hepatitis c
39	11	55.0	137	12	Q914T9	Q914T9 hepatitis c
40	11	55.0	137	12	Q914Q4	Q914Q4 hepatitis c
41	11	55.0	137	12	Q914Q8	Q914Q8 hepatitis c
42	11	55.0	137	12	Q914V1	Q914V1 hepatitis c
43	11	55.0	137	12	Q914V6	Q914V6 hepatitis c
44	11	55.0	137	12	Q914W4	Q914W4 hepatitis c
45	11	55.0	137	12	Q914U6	Q914U6 hepatitis c

ALIGNMENTS

RESULT 1

Q914W1 ID Q914W1 PRELIMINARY; PRT; 137 AA.
AC Q914W1;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Genome polypeptide (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=dog;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before and after liver transplantation."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431826; AAL30676.1; -
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; I.
DR Probom; P186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; Transmembrane.
FT NON TER 1
FT NON TER 137
SQ SEQUENCE 137 AA; 14996 MW; 46F1541F17CFF13F CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 4.3e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYRCGSGPW 20

Db 105 TDCFRKHPDATYRCGSGPW 124

RESULT 2

```
Q81756      PRELIMINARY;      PRT; 2436 AA.
ID Q81756;
AC Q81756;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RA Choo Q.-L., Richman K., Han J.;
RT "The nucleotide sequence of the Hepatitis C viral genome.";
RL Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.
DR EMBL; M32084; AAA45677.1; -.
DR HSSP; P27958; IALV.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002186; HCV NS5b.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA pol DS PS.
DR InterPro; IPR007094; RNA pol PSVir.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01338; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01007; HCV NS4b; 1.
DR Pfam; PF01001; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00507; RDRP POSITIVE; 1.
DR PROSITE; PS05021; RDRP VIRAL; 1.
DR KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
FT NON_TER 1
FT NON_TER 2436
SQ SEQUENCE 2436 AA; 264734 MW; D7B9872900BE3125 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 2436;
Best Local Similarity 100.0%; Pred. No. 5.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDAYSRCGSGPW 20
Db 133 TDCFRKHPDAYSRCGSGPW 152

RESULT 3
Q91FE5      PRELIMINARY;      PRT; 3011 AA.
ID Q91FE5;
AC Q91FE5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RX MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
```

```
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
sequence.";
RL J. Gen. Virol. 82:1291-1297(2001).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
DR EMBL; AF271632; AAF81759.1; -.
DR HSSP; P27958; IALV.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA pol DS PS.
DR InterPro; IPR007094; RNA pol PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
DR PROSITE; PS05057; RDRP POSITIVE; 1.
DR PROSITE; PS05021; RDRP VIRAL; 1.
DR KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327124 MW; 2489CE74AC864E58 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 6.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDAYSRCGSGPW 20
Db 583 TDCFRKHPDAYSRCGSGPW 602

RESULT 4
Q81754      PRELIMINARY;      PRT; 3011 AA.
ID Q81754;
AC Q81754;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RC STRAIN=HC-G9;
RA Okamoto H.;
RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
RN [2]_TaxID=11103;
RP SEQUENCE FROM N.A.
```

RC STRAIN=HC-G9;
 RX MEDLINE=94172337; PubMed=8126459;
 RA Okamoto H., Kojima M., Sakamoto M., Izuka H., Hagiwara S.,
 RA Suwignyo S., Miyakawa Y., Mayumi M.;
 RT "The entire nucleotide sequence and classification of a hepatitis C
 RT virus isolate of a novel genotype from an Indonesian patient with
 RT chronic liver disease.";
 RL J. Gen. Virol. 75:629-635(1994).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL; D14853; BAA03581.1; -.
 DR HSSP; P26663; LUXP.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRP.
 DR InterPro; IPR007095; RNA pol DS PS.
 DR InterPro; IPR007094; RNA pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS05057; RDRP_POSITIVE; 1.
 DR PROSITE; PS0521; RDRP_VIRAL; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 FT CHAIN 1 191
 FT CHAIN 192 383
 FT CHAIN 384 809
 FT CHAIN 810 1006
 FT CHAIN 1007 1657
 FT CHAIN 1658 1972
 FT CHAIN 1973 3011
 SQ SEQUENCE 3011 AA; 327212 MW; 9C16C120F4E79268 CRC64;
 Query Match 100.0%; Score 20; DB 12; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 6.1e-14;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDCFRKHPDATYRCGSGPW 20
 Db 583 TDCFRKHPDATYRCGSGPW 602
 RESULT 5
 Q68870
 ID Q68870 PRELIMINARY; PRT; 3021 AA.
 AC Q68870;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Genes for core, envelope and NS1 proteins (Genome
 DE polyprotein).
 OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11103;
 RN SEQUENCE FROM N.A.
 RA Seelig R., Weber P., Seeling H.P., Ledger N., Bottnar C., Renz M.;
 RA "Hepatitis C virus type V genome isolated from a patient in Germany.";
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL; X76918; CAA54244.1; -.
 DR HSSP; P27958; 1A1V.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRP.
 DR InterPro; IPR007095; RNA pol DS PS.
 DR InterPro; IPR007094; RNA pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS05057; RDRP_POSITIVE; 1.
 DR PROSITE; PS0521; RDRP_VIRAL; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 FT CHAIN 1 191
 FT CHAIN 192 383
 FT CHAIN 384 735
 FT CHAIN 736 1012
 FT CHAIN 1013 1663
 FT CHAIN 1664 1717
 FT CHAIN 1718 1978
 FT CHAIN 1979 2430
 FT CHAIN 2431 3021
 SQ SEQUENCE 3021 AA; 329092 MW; BF2B499AA5A586CF CRC64;
 Query Match 80.0%; Score 16; DB 12; Length 3021;
 Best Local Similarity 100.0%; Pred. No. 2e-09;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDCFRKHPDATYRCG 16
 Db 589 TDCFRKHPDATYRCG 604
 RESULT 6
 Q914S8
 ID Q914S8 PRELIMINARY; PRT; 137 AA.
 AC Q914S8;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).

```

OS Hepatitis C virus
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=nf;
RA Lyra A.C.; Fan X.; Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
RT and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431860; AAL30709.1; -
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 15114 MW; 3A96B0A512453925 CRC64;

Query Match 65.0%; Score 13; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 3.4e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATYS 13
DB 105 TDCFRKHPDATYS 117

RESULT 7
ID Q81323 PRELIMINARY; PRT; 350 AA.
AC Q81323;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2/NS1 protein (Genome polyprotein) (Fragment).
GN E2/NS1.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=T8;
RA Nakao H.; Tsukada K.; Takeuchi T.; Suzuki S.; Hirose A.; Iwano K.;
RA Mitsui T.; Masuko K.; Nakajima S.; Nakajima K.;
RT "Evolutionary analysis of the hepatitis C virus E2/NS1 gene in a
RT patient for 8 years.";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D30727; BAA06392.1; -
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 350
SQ SEQUENCE 350 AA; 39489 MW; DB5CC6B8A10BF57A CRC64;

Query Match 60.0%; Score 12; DB 12; Length 350;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATY 12
DB 204 TDCFRKHPDATY 215

RESULT 8
ID Q81318 PRELIMINARY; PRT; 350 AA.
AC Q81318;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2/NS1 protein (Genome polyprotein) (Fragment).
GN E2/NS1.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=T9;
RA Nakao H.; Tsukada K.; Takeuchi T.; Suzuki S.; Hirose A.; Iwano K.;
RA Mitsui T.; Masuko K.; Nakajima S.; Nakajima K.;
RT "Evolutionary analysis of the hepatitis C virus E2/NS1 gene in a
RT patient for 8 years.";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D30727; BAA06392.1; -
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 350
SQ SEQUENCE 350 AA; 39323 MW; 097724119387A450 CRC64;

Query Match 60.0%; Score 12; DB 12; Length 350;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATY 12
DB 204 TDCFRKHPDATY 215

RESULT 9
ID Q81324 PRELIMINARY; PRT; 350 AA.
AC Q81324;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2/NS1 protein (Genome polyprotein) (Fragment).
GN E2/NS1.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=T9;
RA Nakao H.; Tsukada K.; Takeuchi T.; Suzuki S.; Hirose A.; Iwano K.;
RA Mitsui T.; Masuko K.; Nakajima S.; Nakajima K.;
RT "Evolutionary analysis of the hepatitis C virus E2/NS1 gene in a
RT patient for 8 years.";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D30728; BAA06393.1; -
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 350
SQ SEQUENCE 350 AA; 39323 MW; 097724119387A450 CRC64;

Query Match 60.0%; Score 12; DB 12; Length 350;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATY 12
DB 204 TDCFRKHPDATY 215

```



```
Db          204 TDCFRKHPDATY 215
|||||
RESULT 10
Q81320
ID Q81320 PRELIMINARY; PRT; 350 AA.
AC Q81320;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2/NS1 protein (Genome polyprotein) (Fragment).
CN E2/NS1.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RC STRAIN=T5;
RA Nakao H., Tsukada K., Takeuchi T., Suzuki S., Hirose A., Iwano K.,
RA Mitui T., Masuko K., Nakajima S., Nakajima K.;
RT "Evolutionary analysis of the hepatitis C virus E2/NS1 gene in a
RT patient for 8 years.";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D30724; BAA06389.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 350
SQ SEQUENCE 350 AA; 39485 MW; 94361B69A08359FB CRC64;

Query Match 60.0%; Score 12; DB 12; Length 350;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATY 12
|||||
Db          204 TDCFRKHPDATY 215
|||||
RESULT 11
Q81322
ID Q81322 PRELIMINARY; PRT; 350 AA.
AC Q81322;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2/NS1 protein (Genome polyprotein) (Fragment).
CN E2/NS1.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RC STRAIN=T7;
RA Nakao H., Tsukada K., Takeuchi T., Suzuki S., Hirose A., Iwano K.,
RA Mitui T., Masuko K., Nakajima S., Nakajima K.;
RT "Evolutionary analysis of the hepatitis C virus E2/NS1 gene in a
RT patient for 8 years.";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D30726; BAA06391.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 350
SQ SEQUENCE 350 AA; 39402 MW; 8659E93A8752FE249 CRC64;

Query Match 60.0%; Score 12; DB 12; Length 350;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATY 12
|||||
Db          204 TDCFRKHPDATY 215
|||||
RESULT 13
Q81321
ID Q81321 PRELIMINARY; PRT; 350 AA.
AC Q81321;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2/NS1 protein (Genome polyprotein) (Fragment).
CN E2/NS1.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RC STRAIN=T6;
RA Nakao H., Tsukada K., Takeuchi T., Suzuki S., Hirose A., Iwano K.,
RA Mitui T., Masuko K., Nakajima S., Nakajima K.;
RT "Evolutionary analysis of the hepatitis C virus E2/NS1 gene in a
RT patient for 8 years.";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D30725; BAA06390.1; -.

```

```
SQ SEQUENCE 350 AA; 39502 MW; F3CFE3FCF1049EE2 CRC64;

Query Match 60.0%; Score 12; DB 12; Length 350;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATY 12
|||||
Db          204 TDCFRKHPDATY 215
|||||
RESULT 12
Q81325
ID Q81325 PRELIMINARY; PRT; 350 AA.
AC Q81325;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2/NS1 protein (Genome polyprotein) (Fragment).
CN E2/NS1.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RC STRAIN=T10;
RA Nakao H., Tsukada K., Takeuchi T., Suzuki S., Hirose A., Iwano K.,
RA Mitui T., Masuko K., Nakajima S., Nakajima K.;
RT "Evolutionary analysis of the hepatitis C virus E2/NS1 gene in a
RT patient for 8 years.";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D30729; BAA06394.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 350
SQ SEQUENCE 350 AA; 39402 MW; 8659E93A8752FE249 CRC64;

Query Match 60.0%; Score 12; DB 12; Length 350;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATY 12
|||||
Db          204 TDCFRKHPDATY 215
|||||
RESULT 13
Q81321
ID Q81321 PRELIMINARY; PRT; 350 AA.
AC Q81321;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2/NS1 protein (Genome polyprotein) (Fragment).
CN E2/NS1.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RC STRAIN=T6;
RA Nakao H., Tsukada K., Takeuchi T., Suzuki S., Hirose A., Iwano K.,
RA Mitui T., Masuko K., Nakajima S., Nakajima K.;
RT "Evolutionary analysis of the hepatitis C virus E2/NS1 gene in a
RT patient for 8 years.";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D30725; BAA06390.1; -.

```

```

DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 350
SQ SEQUENCE 350 AA; 39568 MW; 2BC79C74DBDD956 CRC64;

Query Match 60.0%; Score 12; DB 12; Length 350;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATY 12
Db 204 TDCFRKHPDATY 215
|||||

RESULT 14
Q81316 PRELIMINARY; PRT; 350 AA.
AC Q81316;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2/NS1 protein (Genome polyprotein) (Fragment).
GN E2/NS1.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=T1;
RA Nakao H., Tsukada K., Takeuchi T., Suzuki S., Hirose A., Iwano K.,
RA Mitsui T., Masuko K., Nakajima S., Nakajima K.;
RT "Evolutionary analysis of the hepatitis C virus E2/NS1 gene in a
RT patient for 8 years.";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D30721; BAA06386.1; -
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 350
SQ SEQUENCE 350 AA; 39313 MW; 61BAB4B9B41A6AEA CRC64;

Query Match 60.0%; Score 12; DB 12; Length 350;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATY 12
Db 204 TDCFRKHPDATY 215
|||||

Search completed: November 21, 2003, 21:08:20
Job time : 24.45 secs

RC STRAIN=T2;
RA Nakao H., Tsukada K., Takeuchi T., Suzuki S., Hirose A., Iwano K.,
RA Mitsui T., Masuko K., Nakajima S., Nakajima K.;
RT "Evolutionary analysis of the hepatitis C virus E2/NS1 gene in a
RT patient for 8 years.";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D30721; BAA06386.1; -
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 350
SQ SEQUENCE 350 AA; 39313 MW; 61BAB4B9B41A6AEA CRC64;

Query Match 60.0%; Score 12; DB 12; Length 350;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATY 12
Db 204 TDCFRKHPDATY 215
|||||

Query Match 60.0%; Score 12; DB 12; Length 350;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDCFRKHPDATY 12
Db 204 TDCFRKHPDATY 215
|||||

RESULT 15
Q81317 PRELIMINARY; PRT; 350 AA.
AC Q81317;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2/NS1 protein (Genome polyprotein) (Fragment).
GN E2/NS1.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:22 ; Search time 32.1 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-87
Perfect score: 20
Sequence: 1 SRCGGPWITPRLCLVDYPR 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 159726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_13Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	16 AAR84492	Hepatitis C virus
2	20	100.0	20	17 AAR91010	HCV E2 peptide E2-
3	20	100.0	20	23 AAO18716	Hepatitis C virus
4	20	100.0	192	19 AAW67009	HCV nucleocapsid c
5	20	100.0	254	22 AAB68043	Amino acid sequenc
6	20	100.0	363	22 AAB68042	Amino acid sequenc
7	20	100.0	402	14 AAR34438	Sequence of glycop
8	20	100.0	414	14 AAR33589	HCV CKS-NS1S2 fusi
9	20	100.0	414	22 AAB69008	HCV recombinant an

10	20	100.0	454	10 AAP90183	Sequence of hepati
11	20	100.0	454	10 AAP92049	Sequence encoded b
12	20	100.0	454	21 AAB18526	Protein encoded by
13	20	100.0	480	14 AAR33992	HCV-1 E2/NS1 prote
14	20	100.0	531	22 AAE02622	Chimeric HCV E2661
15	20	100.0	621	14 AAR33185	Sequence of subfra
16	20	100.0	622	14 AAR33591	HCV CKS-NS1S1-NS1S
17	20	100.0	622	22 AAB69010	HCV recombinant an
18	20	100.0	637	24 ABP57410	Hepatitis C virus
19	20	100.0	663	17 AAR92935	HCV1 E2 + NS2 poly
20	20	100.0	663	20 AAW67615	Hepatitis C virus
21	20	100.0	733	14 AAR38278	NANB hepatitis vir
22	20	100.0	738	14 AAR33592	HCV CKS-full lengt
23	20	100.0	738	22 AAB69011	HCV recombinant an
24	20	100.0	2435	13 AAR25135	HCV polypeptide 1.
25	20	100.0	2436	10 AAP92050	Sequence encoded i
26	20	100.0	2436	10 AAP90288	Peptide encoded by
27	20	100.0	2436	13 AAR28582	HCV amino acid seq
28	20	100.0	2772	11 AAR08123	Hepatitis C virus
29	20	100.0	2772	21 AAB18540	Protein encoded by
30	20	100.0	2816	14 AAR34009	HCV-1 polyprotein.
31	20	100.0	2894	13 AAR24440	Composite HCV HC-J
32	20	100.0	2894	16 AAR70230	Composite hepatitis
33	20	100.0	2955	20 AAY14975	Amino acid sequenc
34	20	100.0	2955	21 AAB18541	Polyprotein encode
35	20	100.0	3011	13 AAR21519	Compiled HCV sequ
36	20	100.0	3011	15 AAR66995	Hepatitis C virus
37	20	100.0	3011	17 AAR90931	Hepatitis C virus
38	20	100.0	3011	18 AAW34480	HCV polyprotein.
39	20	100.0	3011	19 AAW40038	HCV polyprotein.
40	20	100.0	3011	23 AAE22049	Hepatitis C virus
41	20	100.0	3011	23 AAU84597	HCV polyprotein 1a
42	19	95.0	30	23 AAU84637	HCV HepC1a segment
43	19	95.0	1997	23 AAU84802	HCV HepC1a casette
44	19	95.0	5985	23 AAU84799	HCV HepC1a scrambl
45	16	80.0	30	23 AAU84638	HCV HepC1a segment

ALIGNMENTS

RESULT 1

- AAR84492
- ID AAR84492 standard; peptide; 20 AA.
- XX AC
- XX AC AAR84492;
- XX DT 06-JAN-1997 (first entry)
- XX DE Hepatitis C virus T-cell epitope peptide NSI-25 (residues 595-614).
- XX KW Hepatitis C virus; HCV; immunogen; non-structural region;
- XX KW immunodominant; T cell epitope; vaccine.
- XX OS Hepatitis C virus.
- XX PN WO9512677-A2.
- XX PD 11-MAY-1995.
- XX PF 28-OCT-1994; 94WO-EP03555.
- XX PR 04-NOV-1993; 93EP-0402718.
- XX (INNO-) INNOGENETICS NV.
- XX PA Deleys R, Leroux-Roels G, Maertens G;
- XX PI WPI; 1995-193822/25.
- XX DR Hepatitis C Virus immunogenic polypeptide contg. a T-cell
- XX PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in
- XX PT production of vaccines, therapeutic agents, etc.

XX PS Claim 26; Page 71; 105pp; English.
 XX CC The present sequence is a specifically claimed example of a
 CC T-cell epitope-containing peptide derived from hepatitis C virus.
 XX SQ Sequence 20 AA;
 Query Match 100.0%; Score 20; DB 16; Length 20;
 Best Local Similarity 100.0%; Pred. No. 5.3e-14;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SRCGSGPWITPRCLVDYPYR 20
 |||||
 Db 1 SRCGSGPWITPRCLVDYPYR 20
 |||||
 RESULT 2
 AAR91010
 ID AAR91010 standard; peptide; 20 AA.
 XX AC AAR91010;
 XX DT 25-SEP-1996 (first entry)
 XX DE HCV E2 peptide E2-25 for competition studies.
 XX KW HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;
 KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.
 XX OS Synthetic.
 XX PN WO9604385-A2.
 XX PD 15-FEB-1996.
 XX PF 31-JUL-1995; 95WO-EP03031.
 XX PR 29-JUL-1994; 94EP-0870132.
 XX PA (INNO-) INNOGENETICS NV.
 XX PI Bosman F, Buyse M, De Martynoff G, Maertens G;
 XX WPI; 1996-129401/13.
 XX PT Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope
 PT proteins - in presence of disulphide bond cleavage agent, to
 PT produce proteins suitable for direct use in vaccines or diagnostic
 PT assays of HCV
 XX PS Claim 29; Page 67; 146pp; English.
 XX CC AAR90965-R90980 and AAR90995-R91015 represent synthetic hepatitis C
 CC virus (HCV) E1 and E2 peptides used in competition studies. This
 CC sequence represents a synthetic E2 peptide, and corresponds to residues
 CC 595-614 of the E2 protein sequence. These sequences are useful for in
 CC vitro monitoring of HCV disease, or prognosis of the response to
 CC interferon treatment of patients suffering from HCV infection. These
 CC sequences compete with the proteins produced by AAT12704-T12709 and
 CC AAT12961-T12974, which are included in vectors for the production of
 CC recombinant E1, E2, and E1/E2. The recombinant proteins can then be
 CC isolated and purified by carrying out a disulphide bond cleavage, or a
 CC reduction step with a disulphide bond cleavage agent, after lysis of
 CC recombinant host cells. The constructs containing the purified HCV
 CC envelope proteins can be used for vaccinating humans against HCV, for in
 CC vitro detection of HCV antibodies in a sample, and in a serotyping assay
 CC for detecting one or more serological types of HCV present in a
 CC biological sample. The constructs can also be immobilised on a solid
 CC substrate and incorporated into a reversed phase hybridisation assay for
 CC determining the presence or the genotype of HCV. The new purification
 CC method preserves the conformation of the recombinantly expressed E1, E2
 CC and E1/E2, and eliminates contaminating proteins. Antigens isolated

CC using this method are more reactive with human sera than those isolated
 CC by known techniques.
 XX SQ Sequence 20 AA;
 Query Match 100.0%; Score 20; DB 17; Length 20;
 Best Local Similarity 100.0%; Pred. No. 5.3e-14;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SRCGSGPWITPRCLVDYPYR 20
 |||||
 Db 1 SRCGSGPWITPRCLVDYPYR 20
 |||||
 RESULT 3
 AAO18716
 ID AAO18716 standard; Peptide; 20 AA.
 XX AC AAO18716;
 XX DT 24-OCT-2002 (first entry)
 XX DE Hepatitis C virus E2 protein derived peptide E2-25.
 XX KW Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide;
 KW immunostimulant; vaccine.
 XX OS Hepatitis C virus.
 XX PN WO200255548-A2.
 XX PD 18-JUL-2002.
 XX PF 11-JAN-2002; 2002WO-EP00219.
 XX PR 11-JAN-2001; 2001US-260699P.
 XX PR 30-AUG-2001; 2001US-315768P.
 XX PA (INNO-) INNOGENETICS NV.
 XX PI Maertens G, Bosman F, Buyse M;
 XX WPI; 2002-599657/64.
 XX DR New therapeutic vaccine compositions comprising at least one purified
 PT recombinant hepatitis C virus (HCV) single or specific oligomeric
 PT recombinant envelope protein E1 or E2, useful for immunizing humans
 PT from HCV infection
 XX PS Claim 4; Page 230; 243pp; English.
 XX CC The present invention relates to new therapeutic vaccine compositions for
 CC inducing hepatitis C virus (HCV)-specific antibodies, comprising a
 CC composition containing at least one purified recombinant HCV single or
 CC specific oligomeric recombinant envelope proteins selected from an E1 and
 CC an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are
 CC useful for inducing HCV-specific antibodies or for immunising humans
 CC against HCV. The recombinant HCV E1 and/or E2 proteins are useful as
 CC vaccines or therapeutics, in HCV screening and confirmatory antibody
 CC tests, for raising antibodies, in the preparation of medicament, and for
 CC in vitro monitoring of HCV disease or prognosing the response to
 CC treatment of patients suffering from HCV infection. The present sequence
 CC is a peptide derived from the proteins of the invention.
 XX SQ Sequence 20 AA;
 Query Match 100.0%; Score 20; DB 23; Length 20;
 Best Local Similarity 100.0%; Pred. No. 5.3e-14;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SRCGSGPWITPRCLVDYPYR 20
 |||||
 Db 1 SRCGSGPWITPRCLVDYPYR 20
 |||||

RESULT 4
AAW67009
ID AAW67009 standard; protein; 192 AA.
XX
XX
AC AAW67009;
XX
XX
DT 02-MAR-1999 (first entry)
XX
XX
DE HCV nucleocapsid core protein.
XX
XX
KW Hepatitis C virus; HCV; nucleocapsid; core protein; analogue; antibody;
XX non-structural protein; thioamide bond; peptide bond.
XX
OS Hepatitis C virus.
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 79 /note= "given in specification as Try"
XX
XX
FN JP10226698-A.
XX
XX
PD 25-AUG-1998.
XX
XX
PF 19-FEB-1997; 97JP-0034702.
XX
XX
PR 19-FEB-1997; 97JP-0034702.
XX
XX
PA (KYOW) KYOWA MEDEX KK.
XX
XX
DR WPI; 1998-515103/44.
XX
XX
PT Determination of antibody in sample - uses peptide analog absorbed
XX or chemically bound on carrier as antigen
XX
XX
PS Disclosure; Page 4; 13pp; Japanese.
XX
XX
CC This sequence represents the Hepatitis C virus (HCV) nucleocapsid core
CC protein. The invention relates to peptide analogues derived from HCV
CC proteins, e.g. AAW67417-W67426, which can be used for the determination
CC of anti-HCV antibodies in a sample. Preferably the peptide analogues
CC contain one or more thioamide peptide bonds where at least one oxygen
CC atom of the peptide bond is replaced by sulphur atom. The peptide
XX analogues can be adsorbed or chemically bound to a carrier.
XX
SQ Sequence 192 AA;
Query Match 100.0%; Score 20; DB 19; Length 192;
Best Local Similarity 100.0%; Pred. No. 3.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SRCGSGPWITPRLCLVDYPYR 20
DB 145 SRCGSGPWITPRLCLVDYPYR 164
RESULT 5
AAB68043
ID AAB68043 standard; protein; 254 AA.
XX
XX
AC AAB68043;
XX
XX
DT 29-JUN-2001 (first entry)
XX
XX
DE Amino acid sequence of water soluble variant of envelope E2 protein.
XX
XX
KW Eo protein; HCV; envelope protein; E2 protein; HCV infection;
XX HCV attachment.
XX
XX
OS Synthetic.
XX Hepatitis C virus.
XX

PN WO200122984-A1.
XX
XX
PD 05-APR-2001.
XX
XX
PF 26-SEP-2000; 2000WO-US26395.
XX
XX
PR 29-SEP-1999; 99US-0407430.
XX
XX
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
XX
PI Worman HJ, Mamiya N;
XX
XX
DR WPI; 2001-273486/28.
XX
XX
PT Treating or preventing hepatitis C virus infection in a subject,
XX involves administering hepatitis C virus envelope protein E2 binding
XX agents
XX
XX
PS Claim 5; Fig 8; 46pp; English.
XX
XX
CC The present sequence represents a water soluble variant of a Hepatitis C
CC virus (HCV) envelope E2 protein. Agents that bind to the HCV E2 protein
CC (such as the human Eo protein), and so inhibit the attachment of HCV onto
CC cells (especially liver cells), are used to treat HCV infections in
CC mammals, in particular humans. The specification also describes a method
CC for identifying a compound which can be used for treating or preventing
CC HCV in a subject and which can inhibit the attachment of HCV onto cells
CC by inhibiting the binding of HCV envelope E2 protein to a cellular
CC protein associated with HCV attachment and entry into cells. The method
CC comprises incubating the compound, HCV envelope E2 protein or its variant
CC and a cellular protein capable of specifically binding to the HCV E2
CC protein under suitable reaction conditions; determining the interactions
CC between HCV envelope E2 protein and cellular protein in the presence and
CC absence of the compound; and comparing the interaction to identify a
CC compound which can inhibit the attachment of HCV onto cells.
XX
XX
SQ Sequence 254 AA;
Query Match 100.0%; Score 20; DB 22; Length 254;
Best Local Similarity 100.0%; Pred. No. 4.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SRCGSGPWITPRLCLVDYPYR 20
DB 190 SRCGSGPWITPRLCLVDYPYR 209
RESULT 6
AAB68042
ID AAB68042 standard; protein; 363 AA.
XX
XX
AC AAB68042;
XX
XX
DT 29-JUN-2001 (first entry)
XX
XX
DE Amino acid sequence of a Hepatitis C virus envelope E2 protein.
XX
XX
KW Eo protein; HCV; envelope protein; E2 protein; HCV infection;
XX HCV attachment.
XX
XX
OS Hepatitis C virus.
XX
XX
PN WO200122984-A1.
XX
XX
PD 05-APR-2001.
XX
XX
PF 26-SEP-2000; 2000WO-US26395.
XX
XX
PR 29-SEP-1999; 99US-0407430.
XX
XX
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
XX
PI Worman HJ, Mamiya N;

XX		WPI; 2001-273486/28.
XX		Treating or preventing hepatitis C virus infection in a subject,
XX		involves administering hepatitis C virus envelope protein E2 binding
PT	-	agents -
PT		
XX		Claim 3; Fig 7; 46pp; English.
PS		
XX		The present sequence represents a Hepatitis C virus (HCV) envelope
XX		E2 protein. Agents that bind to the HCV E2 protein (such as the human
CC		Eo protein), and so inhibit the attachment of HCV onto cells
CC		(especially liver cells), are used to treat HCV infections in mammals,
CC		in particular humans. The specification also describes a method for
CC		identifying a compound which can be used for treating or preventing
CC		HCV in a subject and which can inhibit the attachment of HCV onto cells
CC		by inhibiting the binding of HCV envelope E2 protein to a cellular
CC		protein associated with HCV attachment and entry into cells. The method
CC		comprises incubating the compound, HCV envelope E2 protein or its variant
CC		and a cellular protein capable of specifically binding to the HCV E2
CC		protein under suitable reaction conditions; determining the interactions
CC		between HCV envelope E2 protein and cellular protein in the presence and
CC		absence of the compound; and comparing the interaction to identify a
CC		compound which can inhibit the attachment of HCV onto cells.
XX		
XX		Sequence 363 AA;
SQ		
	Query Match	100.0%; Score 20; DB 22; Length 363;
	Best Local Similarity	100.0%; Pred. No. 6e-13;
	Matches 20; Conservative	0; Mismatches 0; Indels 0; Gaps
		0;
Qy	1 SRCGSGFWITPRCLVDYPYR 20	
Dd	212 SRCGSGFWITPRCLVDYPYR 231	
RESULT 7		
AAR34438		
ID	AAR34438 standard; Protein; 402 AA.	
XX		
AC	AAR34438;	
XX		
DT	25-MAR-2003 (updated)	
DT	09-AUG-1993 (first entry)	
XX		
DE	Sequence of glycoprotein E2/NS1 in clone HCV1.	
XX		
KW	Hepatitis C virus; envelope protein; glycoprotien; E2/NS1;	
KW	diagnostic reagent.	
XX		
OS	Hepatitis C virus.	
XX		
PN	EP537626-A1.	
XX		
PD	21-APR-1993.	
XX		
PF	08-OCT-1992; 92EP-0117191.	
XX		
PR	08-OCT-1991; 91JP-0260824.	
XX		
PA	(NAHE-) NAT INST OF HEALTH.	
XX		
Pi	Harada S, Honda Y, Miyamura T, Saito I;	
XX		
DR	WPI; 1993-127516/16.	
DR	N-PSDB; AAO40330.	
XX		
PT	Diagnostic reagent for hepatitis C virus - comprises second	
PT	envelope protein or first non-structural protein encoded by HCV	
PT	gene and has sugar chain	
XX		
PS	Claim 2; Pages 30-32; 58pp; English.	
XX		

Query Match 100.0%; Score 20; DB 14; Length 414;
 Best Local Similarity 100.0%; Pred. No. 6.8e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
 |||||
 Db 278 SRCGSGPWITPRCLVDYPYR 297

RESULT 9

AAB69008
 ID AAB69008 standard; Peptide; 414 AA.

XX AC AAB69008;

XX DT 17-APR-2001 (first entry)

XX DE HCV recombinant antigen pHCV-65 amino acid sequence SEQ ID NO:32.

XX KW Hepatitis C virus; HCV; antigen; detection; antibody.

XX OS Hepatitis C virus.

XX PN US6172189-B1.

XX PD 09-JAN-2001.

XX PF 02-JUN-1997; 97US-0867611.

XX PR 19-NOV-1992; 92US-0989843.

XX PR 10-JAN-1994; 94US-0179896.

XX PR 01-MAY-1996; 96US-0646757.

XX PR 24-AUG-1990; 90US-0572822.

XX PR 07-NOV-1990; 90US-0614069.

XX PR 21-AUG-1991; 91US-0748561.

XX PR 21-AUG-1991; 91US-0748566.

XX PR 29-OCT-1991; 91US-0748565.

XX PA (ABBO) ABBOTT LAB.

XX PI Devare SG, Desai SM, Casey JM, Dailey SH, Dawson GJ, Gutierrez RA;

XX PI Lesniewski RR, Stewart JL, Rupprecht KR;

XX WPI; 2001-122352/13.

XX PT New recombinant antigens representing distinct antigenic regions of
 PT Hepatitis C virus (HCV) genome, useful for detection of antibodies and
 PT antigens in body fluids of individuals exposed to HCV -

XX PS Example 11; Column 141-144; 167pp; English.

XX CC The present invention describes recombinant Hepatitis C virus (HCV)
 CC antigens (I). (I) is useful as a reagent for the detection of antibodies
 CC and antigen in body fluids from individuals exposed to HCV. The HCV
 CC assay uses reliable and efficient reagents and methods to accurately
 CC detect the presence of HCV antibodies in samples obtained from
 CC individuals suspected of having HCV infection. AAF32218 to AAF32235.
 CC AAB51371 to AAB51379 and AAB69001 to AAB69032 represent sequences used
 CC in the exemplification of the present invention.

XX SQ Sequence 414 AA;

Query Match 100.0%; Score 20; DB 22; Length 414;
 Best Local Similarity 100.0%; Pred. No. 6.8e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
 |||||
 Db 278 SRCGSGPWITPRCLVDYPYR 297

RESULT 10

AAP90183

ID AAP90183 standard; protein; 454 AA.
 XX AC AAP90183;

XX DT 25-MAR-2003 (updated)
 XX DT 01-NOV-1989 (first entry)

XX DE Sequence of hepatitis C virus cDNA insert in clone k9-1.

XX KW Hepatitis C virus; clone k9-1; probe; vaccine.

XX OS Pan troglodytes.

XX FH Key Location/Qualifiers
 XX FT Region 97..454

XX PN GB2212511-A.

XX PD 26-JUL-1989.

XX PF 18-NOV-1988; 88GB-0027024.

XX PR 18-NOV-1987; 87US-0122714.

XX PR 30-DEC-1987; 87US-0139886.

XX PR 26-FEB-1988; 88US-0161072.

XX PR 26-OCT-1988; 88US-0263584.

XX PA (CHIR) CHIRON CORPORATION.

XX PI Houghton M, Choo QL, Kuo G;

XX DR WPI; 1989-215054/30.

XX DR N-PSDB; AAN90335.

XX PT Hepatitis C virus gene - used for prodn. of polynucleotide probes,
 PT polypeptide(s) and antibodies for diagnosis, prevention and
 PT treatment of infection.

XX PS Disclosure; fig 46; 235pp; English.

XX CC The sequence is the peptide encoded by the hepatitis C virus
 CC (HCV) cDNA insert in clone k9-1 (see AAN90335). The polypeptides
 CC are used to diagnose HCV-induced NANBH, to raise antibodies for
 CC immunoassay or treatment, or to produce vaccines.
 CC The region shown overlaps the cDNA of AAN90327.
 CC (Updated on 25-MAR-2003 to correct PR field.)

XX SQ Sequence 454 AA;

Query Match 100.0%; Score 20; DB 10; Length 454;
 Best Local Similarity 100.0%; Pred. No. 7.3e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
 |||||
 Db 145 SRCGSGPWITPRCLVDYPYR 164

RESULT 11

AAP92049
 ID AAP92049 standard; protein; 454 AA.

XX AC AAP92049;

XX DT 25-MAR-2003 (updated)

XX DT 21-JAN-1991 (first entry)

XX DE Sequence encoded by segment of the hepatitis C virus (HCV) cDNA sequence
 XX DE in clone K9-1.

XX KW Non-a non-B hepatitis; probe; vaccine; diagnosis;
 XX KW passive immunotherapy; antigen.

OS Hepatitis C virus.

XX EP318216-A.
 XX 31-MAY-1989.
 XX 18-NOV-1988; 88EP-0310922.
 XX 18-NOV-1987; 87US-0122714.
 PR 30-DEC-1987; 87US-0139886.
 PR 26-FEB-1988; 88US-0161072.
 PR 06-MAY-1988; 88US-0191263.
 PR 26-OCT-1988; 88US-0263584.
 PR 14-NOV-1988; 88US-0271450.

XX (CHIR) CHIRON CORP.

XX Houghton M, Choo QL, Kuo G;

XX WPI; 1989-159274/22.
 DR N-PSDB; AAN92105.

XX Purified hepatitis C virus -
 PT and assorted nucleic acids and polypeptide(s)

XX Example; Fig 46-1 - 46-2; 139pp; English.

XX Purified hepatitis C virus (HCV) and purified or recombinant HCV nucleic
 CC acids (NAS), encoding HCV polynucleotides or epitopes, and polypeptides
 CC are claimed. HCV is a causative agent of non-A, non-B hepatitis (NANBH).
 CC The NAS may be used to design probes for detn. of HCV NAS in samples.
 CC The polypeptides may be used as immunoassay reagents and vaccines, and
 CC to produce antibodies useful for diagnosis and passive immunotherapy.
 CC The purified virus may also be used in vaccines.

CC (Updated on 25-MAR-2003 to correct PR field.)
 CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 454 AA;

Query Match 100.0%; Score 20; DB 10; Length 454;

Best Local Similarity 100.0%; Pred. No. 7.3e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRCGSGPWITPRLVDYPYR 20

DB 145 SRCGSGPWITPRLVDYPYR 164

RESULT 12

AAB18526

ID AAB18526 standard; Protein; 454 AA.

XX AAB18526;

XX 15-JAN-2001 (first entry)

XX Protein encoded by a novel hepatitis C virus cDNA clone k9-1.

XX Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
 KW viral infectivity; viral replication.

XX Hepatitis C virus.

XX EP1034785-A2.

XX 13-SEP-2000.

XX 16-MAR-1990; 2000EP-0109602.

XX 17-MAR-1989; 89US-0325338.

PR 20-APR-1989; 89US-0341334.

PR 18-MAY-1989; 89US-0355002.

PR 16-MAR-1990; 90EP-0302866.

XX (CHIR) CHIRON CORP.

XX Houghton M, Choo Q, Kuo G;

XX WPI; 2000-566891/53.
 DR N-PSDB; AAR75282.

XX Novel composition comprising a hepatitis C virus antisense
 PT polynucleotide which is complementary to or corresponds to a sense
 PT strand of the virus genome, and selectively hybridises to it -

XX Example; Fig 2; 75pp; English.

XX The specification describes a pharmaceutical composition which
 CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
 CC HCV is characterized by a positive stranded RNA genome which has
 CC 40% homology at the polypeptide level to a HCV polyprotein. The
 CC antisense polynucleotide binds to cellular polynucleotides which
 CC enhance and/or are required for viral infectivity, replicative
 CC ability or chronicity. The antisense polynucleotides may also be
 CC designed to bind with high specificity, to be of increased stability,
 CC to be stable and to have low toxicity. The composition also comprises
 CC an agent which causes viral RNA to be inactive. The composition
 CC is used for preventing HCV replication in a system. The present
 CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
 CC course of the invention.

XX Sequence 454 AA;

Query Match 100.0%; Score 20; DB 21; Length 454;

Best Local Similarity 100.0%; Pred. No. 7.3e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRCGSGPWITPRLVDYPYR 20

DB 145 SRCGSGPWITPRLVDYPYR 164

RESULT 13

AAR33992

ID AAR33992 standard; Protein; 480 AA.

XX AAR33992;

XX 25-MAR-2003 (updated)

DT 26-JUL-1993 (first entry)

XX HCV-1 E2/NS1 protein.

XX Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
 KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
 KW domain; immunological; cross-reactive; envelope protein; vaccine;
 KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.

XX Synthetic.

XX WO9306126-A1.

XX 01-APR-1993.

XX 11-SEP-1992; 92WO-US07683.

XX 13-SEP-1991; 91US-0759575.

XX (CHIR) CHIRON CORP.

XX Houghton M, Weiner AJ;

XX WPI; 1993-117468/14.

XX Immuno-reactive hepatitis C virus polypeptide compans. - contg.
 PT at least 2 sequences from the first variable domain of distinct

PT HCV isolates
 XX Disclosure; Fig 3; 106pp; English.
 PS
 CC The sequences given in AAR33992-002 represent a portion of the E2/NS1
 CC protein encoded by group I and group II HCV isolates, from amino acid
 CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
 CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
 CC 30 amino acids which shows large variation between nearly all isolates.
 CC This is an important immunoreactive domain. This putative envelope
 CC glycoprotein E2/NS1 may correspond to the gp53(BVNV)/gp55 (hog cholera
 CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
 CC flaviviruses, both of which confer protective immunity in hosts
 CC vaccinated with these polypeptides. It has been discovered that a
 CC number of important HCV epitopes vary among viral isolates and that
 CC these epitopes can be mapped to specific domains. This meant that
 CC immunologically cross-reactive polypeptides which focus on variable
 CC rather than constant domains can be produced. See also AAQ39134-48
 CC and AAR33982-91.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 480 AA;
 Query Match 100.0%; Score 20; DB 14; Length 480;
 Best Local Similarity 100.0%; Pred. No. 7.7e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SRCGSGPWITPRCLVDYPYR 20
 |||||
 DB 226 SRCGSGPWITPRCLVDYPYR 245
 |||||
 RESULT 14
 AAEO2622
 ID AAE02622 standard; Protein; 531 AA.
 XX
 AC AAE02622;
 XX
 DT 06-AUG-2001 (first entry)
 DT
 XX
 DE Chimeric HCV E2661-HBsAg S domain encoded by pCMV-II-E2661-SAg.
 XX
 KW Virus-like particle; immunogen; hepatitis B virus surface antigen;
 KW HBsAg; HBV; chimeric antigen; hepatitis C virus; HCV; virucide; vaccine;
 KW plasmid pCMV-II-E2661-SAg; E2 envelope glycoprotein; S domain.
 XX
 OS Chimeric - Hepatitis C virus.
 OS Chimeric - Hepatitis B virus.
 XX
 FH Key Location/Qualifiers
 FT Region 1..303
 FT /note= "HCV 661 E2 envelope glycoprotein"
 FT Region 306..531
 FT /note= "HBsAg S domain"
 FT
 XX WO200138358-A2.
 XX
 XX 31-MAY-2001.
 XX
 XX 22-NOV-2000; 2000WO-US32249.
 XX
 XX 24-NOV-1999; 99US-0167224.
 XX
 XX (CHIR) CHIRON CORP.
 XX
 XX Selby M, Glazer E, Houghton M;
 XX
 XX WPI; 2001-367661/38.
 DR N-PSDB; AAD06793.
 XX
 XX Virus-like particle for use as an immunogen, comprising a first
 PT hepatitis B virus surface antigen (HBsAg) and chimeric antigen
 PT comprising a second HBsAg covalently linked to hepatitis C immunogenic

PT polypeptide -
 XX Claim 29; Fig 4; 115pp; English.
 PS
 CC The invention relates to a virus-like particle for use as an immunogen,
 CC comprising a first hepatitis B virus surface antigen (HBsAg) and a
 CC chimeric antigen comprising a second HBsAg which is covalently linked to
 CC an hepatitis C virus (HCV) immunogenic polypeptide, where the first and
 CC the second HBsAg each comprise a substantially complete S domain.
 CC The virus-like particle is useful as immunogen and as vaccine.
 CC The present sequence is a chimeric antigen comprising HCV 661 E2
 CC envelope glycoprotein and HBsAg S domain encoded by plasmid
 CC pCMV-II-E2661-SAg.
 XX
 SQ Sequence 531 AA;
 Query Match 100.0%; Score 20; DB 22; Length 531;
 Best Local Similarity 100.0%; Pred. No. 8.3e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SRCGSGPWITPRCLVDYPYR 20
 |||||
 DB 237 SRCGSGPWITPRCLVDYPYR 256
 |||||
 RESULT 15
 AAR33185
 ID AAR33185 standard; peptide; 621 AA.
 XX
 AC AAR33185;
 XX
 DT 25-MAR-2003 (updated)
 DT 03-JUL-1993 (first entry)
 XX
 DE Sequence of subfragment pHCV80 (AAs 365-731) of the hepatitis
 DE C virus (HCV) genome.
 XX
 KW Immunogenic peptide; hepatitis C virus; immunogenic domain;
 KW monoclonal antibody; diagnosis; detection; therapy.
 XX
 OS Synthetic.
 XX
 PN WO9304205-A1.
 XX
 PD 04-MAR-1993.
 XX
 PF 21-AUG-1992; 92WO-US07189.
 XX
 PR 21-AUG-1991; 91US-0748292.
 XX
 XX (ABBO) ABBOTT LAB.
 XX
 XX Dailey SH, Desai SM, Devare SG, Johnson JE, Mehta SU;
 XX WPI; 1993-094030/11.
 XX
 XX Monoclonal antibodies specific for hepatitis C virus E2-NS1
 PT antigen - useful for diagnosis and evaluation of HCV infections
 PT and in differentiation studies
 XX
 XX Example; Pages 31-34; 48pp; English.
 PS
 XX Immunogenic domains of E2/NS1 region of HCV genome encompassing AAs
 CC 600-720 were mapped with PEPCAN analysis. Based on the EIA reactivity
 CC of a panel of HCV positive sera, peptide AAR33184 was chosen as the
 CC immunogen for the generation of monoclonal antibodies to HCV NS1.
 CC Several individual oligos representing AAs 365-731 of HCV genome
 CC were ligated and cloned as three separate EcoRI/BamHI subfragments
 CC into the CKS fusion vector pJ0200. The sequences of these three
 CC subfragments - pHCV80, pHCV77 and pHCV65 are in AAR33185-7.
 CC Analysis showed that MAbs H13C113 and H23C163 showed reactivity
 CC pHCV80 and pHCV65.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX
SQ Sequence 621 AA;
Query Match 100.0%; Score 20; DB 14; Length 621;
Best Local Similarity 100.0%; Pred. NO. 9.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SRCGSGPWITPRCLVDYPYR 20
DB 485 SRCGSGPWITPRCLVDYPYR 504

Search completed: November 21, 2003, 20:58:06
Job time : 32.15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.6 Seconds
(without alignments)
79.832 Million cell updates/sec

Title: US-09-973-025-87

Perfect score: 20

Sequence: 1 SRCGSGPWITPRCLVDYPR 20

Scoring table:

OLIGO Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA*
1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS-COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-08-612-973-87
2	20	100.0	20	4	US-08-927-597-87
3	20	100.0	20	4	US-08-635-886C-43
4	20	100.0	42	1	US-08-262-037-44
5	20	100.0	68	4	US-08-635-886C-250
6	20	100.0	68	4	US-08-635-886C-252
7	20	100.0	88	1	US-08-440-103-26
8	20	100.0	88	1	US-08-440-103-27
9	20	100.0	88	1	US-08-440-542-26
10	20	100.0	88	1	US-08-440-542-27
11	20	100.0	88	1	US-08-231-368-26
12	20	100.0	88	1	US-08-231-368-27
13	20	100.0	88	1	US-08-440-210-26
14	20	100.0	88	1	US-08-440-210-27
15	20	100.0	88	4	US-08-046-604-26
16	20	100.0	88	4	US-09-046-604-27
17	20	100.0	179	3	US-08-444-818-77
18	20	100.0	402	1	US-08-460-806-13
19	20	100.0	402	1	US-08-325-630-13
20	20	100.0	403	2	US-08-483-695-39
21	20	100.0	403	2	US-07-965-285-39
22	20	100.0	403	2	US-08-487-231-39
23	20	100.0	403	3	US-09-201-912-39
24	20	100.0	414	1	US-07-748-292-8
25	20	100.0	414	3	US-08-867-611-32
26	20	100.0	414	5	PCT-US92-06965A-2
27	20	100.0	454	3	US-08-444-818-73

28 20 100.0 480 1 US-08-440-103-14 Sequence 14, Appl
29 20 100.0 480 1 US-08-440-542-14 Sequence 14, Appl
30 20 100.0 480 1 US-08-231-368-14 Sequence 14, Appl
31 20 100.0 480 1 US-08-440-210-14 Sequence 14, Appl
32 20 100.0 480 4 US-09-046-604-14 Sequence 14, Appl
33 20 100.0 621 1 US-07-748-292-7 Sequence 7, Appl
34 20 100.0 622 3 US-08-867-611-34 Sequence 34, Appl
35 20 100.0 622 5 PCT-US92-06965A-4 Sequence 4, Appl
36 20 100.0 663 3 US-08-824-057-3 Sequence 3, Appl
37 20 100.0 663 4 US-09-415-582-3 Sequence 3, Appl
38 20 100.0 663 4 US-09-693-596-4 Sequence 4, Appl
39 20 100.0 738 3 US-08-867-611-35 Sequence 35, Appl
40 20 100.0 738 5 PCT-US92-06965A-5 Sequence 35, Appl
41 20 100.0 2436 3 US-08-444-818-75 Sequence 75, Appl
42 20 100.0 2772 3 US-08-444-818-89 Sequence 89, Appl
43 20 100.0 2894 2 US-08-466-975A-23 Sequence 23, Appl
44 20 100.0 2894 2 US-08-391-671A-23 Sequence 23, Appl
45 20 100.0 2894 3 US-08-467-902A-23 Sequence 23, Appl

ALIGNMENTS

RESULT 1
US-08-612-973-87
; Sequence 87, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P. C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 87:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-612-973-87
Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SRCGSGPWITPRCLVDYPR 20
|||||

Db 1 SRCGSGPWITPRCLVDYPYR 20

RESULT 2

US-08-927-597-87

Sequence 87, Application US/08927597

Patent No. 6245503

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

APPLICANT: BOSMAN, FONS

APPLICANT: DE MARTYNOFF, GUY

APPLICANT: BUYSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHUYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/927,597

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 87:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-927-597-87

Query Match 100.0%; Score 20; DB 3; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.8e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20

Db 1 SRCGSGPWITPRCLVDYPYR 20

RESULT 3

US-08-635-886C-43

Sequence 43, Application US/08635886C

Patent No. 6555114

GENERAL INFORMATION:

APPLICANT: LEROUX-ROELS, Geert

APPLICANT: DELEYS, Robert

APPLICANT: MAERTENS, Geert

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C

TITLE OF INVENTION: VIRUS

FILE REFERENCE: 2752-18

CURRENT APPLICATION NUMBER: US/08/635,886C

CURRENT FILING DATE: 1996-04-25

PRIOR APPLICATION NUMBER: PCT/EP94/03555

US-08-927-597-87

Query Match 100.0%; Score 20; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.8e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20

Db 1 SRCGSGPWITPRCLVDYPYR 20

Query Match 100.0%; Score 20; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.8e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20

Db 1 SRCGSGPWITPRCLVDYPYR 20

RESULT 4

US-08-262-037-44

Sequence 44, Application US/08262037

Patent No. 5747239

GENERAL INFORMATION:

APPLICANT: Chang Yi Wang and Barbara Hosein

TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR

TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV

TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES

NUMBER OF SEQUENCES: 136

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVE.

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/262,037

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/719,819

FILING DATE: 24-June-1991

APPLICATION NUMBER: 07/667,275

FILING DATE: 11-Mar-1991

APPLICATION NUMBER: 07/651,735

FILING DATE: 07-Feb-1991

APPLICATION NUMBER: 07/558,799

FILING DATE: 26-July-1990

APPLICATION NUMBER: 07/510,153

FILING DATE: 16-April-1990

ATTORNEY/AGENT INFORMATION:

NAME: Maria C. H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4043 US3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-758-4800

TELEFAX: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 44:

SEQUENCE CHARACTERISTICS:

LENGTH: 42 amino acids

TYPE: Amino acid

STRANDEDNESS:

TOPOLOGY: Unknown

US-08-262-037-44

Query Match 100.0%; Score 20; DB 1; Length 42;
Best Local Similarity 100.0%; Pred. No. 3.4e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYDYPYR 20
DB 17 SRCGSGPWITPRCLVDYDYPYR 36

RESULT 5
US-08-635-886C-250
; Sequence 250, Application US/08635886C
; Patent No. 655114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 250
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-250

Query Match 100.0%; Score 20; DB 4; Length 68;
Best Local Similarity 100.0%; Pred. No. 5.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYDYPYR 20
DB 25 SRCGSGPWITPRCLVDYDYPYR 44

RESULT 6
US-08-635-886C-252
; Sequence 252, Application US/08635886C
; Patent No. 655114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 252
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-252

Query Match 100.0%; Score 20; DB 4; Length 68;
Best Local Similarity 100.0%; Pred. No. 5.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYDYPYR 20

DB 25 SRCGSGPWITPRCLVDYDYPYR 44

RESULT 7
US-08-440-103-26
; Sequence 26, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,103
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-103-26

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYDYPYR 20
DB 42 SRCGSGPWITPRCLVDYDYPYR 61

RESULT 8
US-08-440-103-27
; Sequence 27, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA

```

; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,103
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-103-27

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
DB 42 SRCGSGPWITPRCLVDYPYR 61

RESULT 9
US-08-440-542-26
; Sequence 26, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-542-26

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
DB 42 SRCGSGPWITPRCLVDYPYR 61

RESULT 9
US-08-440-542-26
; Sequence 26, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
```

```

; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-542-26

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
DB 42 SRCGSGPWITPRCLVDYPYR 61

RESULT 10
US-08-440-542-27
; Sequence 27, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-542-27

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 SRCGSPWITPRCLVDYDYPYR 20
Db 42 SRCGSPWITPRCLVDYDYPYR 61

RESULT 11

US-08-231-368-26
; Sequence 26, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25.
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/231,368
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-231-368-26

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSPWITPRCLVDYDYPYR 20
Db 42 SRCGSPWITPRCLVDYDYPYR 61

RESULT 12

US-08-231-368-27
; Sequence 27, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA

; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/231,368
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-231-368-27

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSPWITPRCLVDYDYPYR 20
Db 42 SRCGSPWITPRCLVDYDYPYR 61

RESULT 13

US-08-440-210-26
; Sequence 26, Application US/08440210
; Patent No. 5766845
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,210
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-210-26

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
|||||
Db 42 SRCGSGPWITPRCLVDYPYR 61

RESULT 14

US-08-440-210-27
Sequence 27, Application US/08440210
Patent No. 5766845

GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,210
FILING DATE: 12-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368

FILING DATE:
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-210-27

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
|||||
Db 42 SRCGSGPWITPRCLVDYPYR 61

RESULT 15

US-09-046-604-26
Sequence 26, Application US/09046604
Patent No. 6303292

GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/046,604
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368

FILING DATE:
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-046-604-26

Query Match 100.0%; Score 20; DB 4; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
|||||
Db 42 SRCGSGPWITPRCLVDYPYR 61

Search completed: November 21, 2003, 21:15:17
Job time: 10.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time: 19.5 Seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-87

Perfect score: 20

Sequence: 1 SRCGSPWTRCLVDYDYPYR 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PTC_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	10	US-09-973-025-87
2	20	100.0	20	11	US-09-899-303-87
3	20	100.0	20	11	US-09-995-808-87
4	20	100.0	20	11	US-09-995-860-87
5	20	100.0	20	12	US-09-995-791-87
6	20	100.0	254	10	US-09-407-430-3
7	20	100.0	363	10	US-09-407-430-2
8	20	100.0	637	12	US-10-187-257-4
9	20	100.0	637	12	US-10-265-083-2
10	20	100.0	2894	10	US-09-941-611-23
11	20	100.0	2894	15	US-10-044-995-23
12	20	100.0	3011	9	US-09-916-359-2
13	13	65.0	20	15	US-10-127-746-17
14	13	65.0	20	15	US-10-150-165-17
15	13	65.0	176	10	US-09-921-397-81

16	13	65.0	250	10	US-09-952-572-8	Sequence 8, Appli
17	13	65.0	363	12	US-10-128-587A-97	Sequence 97, Appl
18	13	65.0	363	15	US-10-128-590-97	Sequence 97, Appl
19	13	65.0	3011	9	US-09-742-659-4	Sequence 4, Appli
20	13	65.0	3011	10	US-09-238-076-20	Sequence 20, Appl
21	13	65.0	3011	10	US-09-952-572-9	Sequence 9, Appli
22	13	65.0	3011	10	US-09-747-419-20	Sequence 20, Appl
23	13	65.0	3011	11	US-09-891-894-3	Sequence 3, Appli
24	13	65.0	3011	11	US-09-995-937-20	Sequence 20, Appl
25	13	65.0	3011	11	US-09-917-563-20	Sequence 20, Appl
26	13	65.0	3011	12	US-10-184-150-3	Sequence 3, Appli
27	13	65.0	3011	15	US-10-259-275-20	Sequence 20, Appl
28	13	65.0	3012	10	US-09-238-076-2	Sequence 2, Appli
29	13	65.0	3012	11	US-09-995-937-2	Sequence 2, Appli
30	13	65.0	3012	11	US-09-917-563-2	Sequence 2, Appli
31	11	55.0	3011	16	US-10-232-643-6	Sequence 6, Appli
32	8	40.0	20	10	US-09-973-025-86	Sequence 86, Appl
33	8	40.0	20	10	US-09-973-025-88	Sequence 88, Appl
34	8	40.0	20	11	US-09-899-303-86	Sequence 86, Appl
35	8	40.0	20	11	US-09-899-303-88	Sequence 88, Appl
36	8	40.0	20	11	US-09-995-808-86	Sequence 86, Appl
37	8	40.0	20	11	US-09-995-808-88	Sequence 88, Appl
38	8	40.0	20	11	US-09-995-860-86	Sequence 86, Appl
39	8	40.0	20	11	US-09-995-860-88	Sequence 88, Appl
40	8	40.0	20	12	US-09-995-791-86	Sequence 86, Appl
41	8	40.0	20	12	US-09-995-791-88	Sequence 88, Appl
42	7	35.0	45	16	US-10-318-200-16	Sequence 16, Appl
43	7	35.0	165	9	US-09-410-194-6	Sequence 6, Appli
44	7	35.0	290	12	US-10-128-587A-3	Sequence 3, Appli
45	7	35.0	290	15	US-10-128-590-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-973-025-87
; Sequence 87, Application US/09973025
; Publication NO. US20020182706A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/0973,025
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100

; INFORMATION FOR SEQ ID NO: 87:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 87:
US-09-973-025-87

Query Match 100.0%; Score 20; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 SRCGSGPWITPRCLVDYPYR 20

RESULT 2
US-09-899-303-87
; Sequence 87, Application US/09899303
; Publication No. US20030036110A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (RPO)
; CURRENT APPLICATION DATA: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 87:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 87:
US-09-899-303-87

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
| | | | | | | | | | | | | | | | | | | | | |

Db 1 SRCGSGPWITPRCLVDYPYR 20

RESULT 3
US-09-995-808-87
; Sequence 87, Application US/09995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 87
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-87

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 SRCGSGPWITPRCLVDYPYR 20

RESULT 4
US-09-995-860-87
; Sequence 87, Application US/09995860
; Publication No. US20030118603A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 87
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-87

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 SRCGSGPWITPRCLVDYPYR 20

RESULT 5
US-09-995-791-87
; Sequence 87, Application US/09995791
; Publication No. US20030147918A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 87

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
| | | | | | | | | | | | | | | | | | | | | |

```
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-87

Query Match      100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRCGSGPWITPRCLVDYPYR 20
|||||
Db 1 SRCGSGPWITPRCLVDYPYR 20
|||||

RESULT 6
US-09-407-430-3
; Sequence 3, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-407-430-3

Query Match      100.0%; Score 20; DB 10; Length 254;
Best Local Similarity 100.0%; Pred. No. 5.7e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRCGSGPWITPRCLVDYPYR 20
|||||
Db 190 SRCGSGPWITPRCLVDYPYR 209
|||||

RESULT 7
US-09-407-430-2
; Sequence 2, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-09-407-430-2

Query Match      100.0%; Score 20; DB 10; Length 363;
Best Local Similarity 100.0%; Pred. No. 7.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRCGSGPWITPRCLVDYPYR 20
|||||
Db 212 SRCGSGPWITPRCLVDYPYR 231
|||||

RESULT 8

US-10-187-257-4
; Sequence 4, Application US/10187257
; Publication No. US20030138458A1
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: O'HAGAN, Derek
; TITLE OF INVENTION: HCV E1E2 VACCINE COMPOSITIONS
; FILE REFERENCE: 2302-17206
; CURRENT APPLICATION NUMBER: US/10/187,257
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
US-10-187-257-4

Query Match      100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 1.1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRCGSGPWITPRCLVDYPYR 20
|||||
Db 423 SRCGSGPWITPRCLVDYPYR 442
|||||

RESULT 9
US-10-265-083-2
; Sequence 2, Application US/10265083
; Publication No. US20030170273A1
; GENERAL INFORMATION:
; APPLICANT: O'HAGAN, Derek
; APPLICANT: VALIANTE, Nicholas
; TITLE OF INVENTION: ADJUVANT COMPOSITIONS
; FILE REFERENCE: 2302-18357.30
; CURRENT APPLICATION NUMBER: US/10/265,083
; CURRENT FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7
US-10-265-083-2

Query Match      100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 1.1e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRCGSGPWITPRCLVDYPYR 20
|||||
Db 423 SRCGSGPWITPRCLVDYPYR 442
|||||

RESULT 10
US-09-941-611-23
; Sequence 23, Application US/09941611
; Patent No. US20020106640A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSWIN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; ANTIBODIES TO HEPATITIS C VIRUS
```

NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/941,611
FILING DATE: 30-AUG-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: 1995-02-21
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-941-611-23

Query Match 100.0%; Score 20; DB 10; Length 2894;
Best Local Similarity 100.0%; Pred. No. 3.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
|||||

Db 595 SRCGSGPWITPRCLVDYPYR 614

RESULT 11
US-10-044-995-23
Sequence 23, Application US/10044995
Publication No. US20030049685A1
GENERAL INFORMATION:
APPLICANT: DELEYS, ROBERT J
POLLET, DIRK
MAERTENS, GEERT
VAN HEUVERSWIJN, HUGO
TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
ANTIBODIES TO HEPATITIS C VIRUS

NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/044,995
FILING DATE: 15-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: <Unknown>
APPLICATION NUMBER: US 07/920,286
FILING DATE: 14-OCT-1992
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-044-995-23

Query Match 100.0%; Score 20; DB 15; Length 2894;
Best Local Similarity 100.0%; Pred. No. 3.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
|||||

Db 595 SRCGSGPWITPRCLVDYPYR 614

RESULT 12
US-09-916-359-2
Sequence 2, Application US/09916359
Patent No. US20020034734A1
GENERAL INFORMATION:
APPLICANT: Veronique Barban
TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
TREATING C HEPATITIS
FILE REFERENCE: PMCF97-03A
CURRENT APPLICATION NUMBER: US/09/916,359
CURRENT FILING DATE: 2001-07-26
PRIOR APPLICATION NUMBER: 09/388,874
PRIOR FILING DATE: 1999-09-02
PRIOR APPLICATION NUMBER: 97/02,887
PRIOR FILING DATE: 1997-03-06
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 3011
TYPE: PRT
ORGANISM: Virus
US-09-916-359-2

Query Match 100.0%; Score 20; DB 9; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.6e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
|||||

Db 595 SRCGSGPWITPRCLVDYDPR 614

RESULT 13

US-10-127-746-17
; Sequence 17, Application US/10127746
; Publication No. US20030091590A1
; GENERAL INFORMATION:
; APPLICANT: Schnell, Matthias J.
; APPLICANT: Pomerantz, Roger J.
; TITLE OF INVENTION: Recombinant Rhabdoviruses as Live Viral
; TITLE OF INVENTION: Vaccines
; FILE REFERENCE: SCH02.CP201
; CURRENT APPLICATION NUMBER: US/10/127,746
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 09/494,262
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 09/761,312
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/285,552
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic peptide
US-10-127-746-17

Query Match 65.0%; Score 13; DB 15; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRCGSGPWITPRC 13
Db 5 SRCGSGPWITPRC 17

RESULT 14

US-10-150-165-17
; Sequence 17, Application US/10150165
; Publication No. US20030124146A1
; GENERAL INFORMATION:
; APPLICANT: Schnell, Matthias J.
; APPLICANT: Pomerantz, Roger J.
; TITLE OF INVENTION: Recombinant Rhabdoviruses as Live Viral
; TITLE OF INVENTION: Vaccines
; FILE REFERENCE: SCH02.PC301
; CURRENT APPLICATION NUMBER: US/10/150,165
; CURRENT FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 09/494,262
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 09/761,312
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: SCH02-CP201
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 60/285,552
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic peptide
US-10-150-165-17

Query Match 65.0%; Score 13; DB 15; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRCGSGPWITPRC 13
Db 5 SRCGSGPWITPRC 17

RESULT 15

US-09-921-397-81
; Sequence 81, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 81
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-81

Query Match 65.0%; Score 13; DB 10; Length 176;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRCGSGPWITPRC 13
Db 127 SRCGSGPWITPRC 139

Search completed: November 21, 2003, 22:19:40
Job time : 19.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 163.1 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-87

Perfect score: 20

Sequence: 1 SRCGSGPWTRCLVDYPR 20

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA_Main:*

- 1: /cgn2_6/ptodata/1/paa/PCTUS COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06 COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07 COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08 COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US081 COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US082 COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US083 COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US084 COMB.pep.*
- 9: /cgn2_6/ptodata/1/paa/US085 COMB.pep.*
- 10: /cgn2_6/ptodata/1/paa/US086 COMB.pep.*
- 11: /cgn2_6/ptodata/1/paa/US087 COMB.pep.*
- 12: /cgn2_6/ptodata/1/paa/US088 COMB.pep.*
- 13: /cgn2_6/ptodata/1/paa/US089 COMB.pep.*
- 14: /cgn2_6/ptodata/1/paa/US090 COMB.pep.*
- 15: /cgn2_6/ptodata/1/paa/US091 COMB.pep.*
- 16: /cgn2_6/ptodata/1/paa/US092 COMB.pep.*
- 17: /cgn2_6/ptodata/1/paa/US093 COMB.pep.*
- 18: /cgn2_6/ptodata/1/paa/US094 COMB.pep.*
- 19: /cgn2_6/ptodata/1/paa/US095 COMB.pep.*
- 20: /cgn2_6/ptodata/1/paa/US096 COMB.pep.*
- 21: /cgn2_6/ptodata/1/paa/US097A COMB.pep.*
- 22: /cgn2_6/ptodata/1/paa/US097B COMB.pep.*
- 23: /cgn2_6/ptodata/1/paa/US098 COMB.pep.*
- 24: /cgn2_6/ptodata/1/paa/US099A COMB.pep.*
- 25: /cgn2_6/ptodata/1/paa/US099B COMB.pep.*
- 26: /cgn2_6/ptodata/1/paa/US100 COMB.pep.*
- 27: /cgn2_6/ptodata/1/paa/US101 COMB.pep.*
- 28: /cgn2_6/ptodata/1/paa/US102 COMB.pep.*
- 29: /cgn2_6/ptodata/1/paa/US103 COMB.pep.*
- 30: /cgn2_6/ptodata/1/paa/US104 COMB.pep.*
- 31: /cgn2_6/ptodata/1/paa/US106 COMB.pep.*
- 32: /cgn2_6/ptodata/1/paa/US60 COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	ID	Description
1	20	100.0	20 13	US-08-974-685-43
				Sequence 43, Appl

2	20	100.0	20 13	US-08-974-690-43	Sequence 43, Appl
3	20	100.0	20 13	US-08-974-690A-43	Sequence 43, Appl
4	20	100.0	20 13	US-08-974-690B-43	Sequence 43, Appl
5	20	100.0	20 13	US-08-974-690C-43	Sequence 43, Appl
6	20	100.0	20 23	US-09-899-303-87	Sequence 87, Appl
7	20	100.0	20 23	US-09-899-303A-87	Sequence 87, Appl
8	20	100.0	20 25	US-09-973-025-87	Sequence 87, Appl
9	20	100.0	20 25	US-09-995-791-87	Sequence 87, Appl
10	20	100.0	20 25	US-09-995-808-87	Sequence 87, Appl
11	20	100.0	20 25	US-09-995-860-87	Sequence 87, Appl
12	20	100.0	20 26	US-10-020-510-87	Sequence 87, Appl
13	20	100.0	20 29	US-10-321-798-87	Sequence 87, Appl
14	20	100.0	42 8	US-08-475-482-44	Sequence 44, Appl
15	20	100.0	42 8	US-08-477-072-44	Sequence 44, Appl
16	20	100.0	42 8	US-08-477-582-44	Sequence 44, Appl
17	20	100.0	42 8	US-08-480-253-44	Sequence 44, Appl
18	20	100.0	68 13	US-08-974-690C-250	Sequence 250, App
19	20	100.0	68 13	US-08-974-690C-252	Sequence 252, App
20	20	100.0	88 8	US-08-471-498-26	Sequence 26, Appl
21	20	100.0	88 8	US-08-471-498-27	Sequence 27, Appl
22	20	100.0	112 8	US-08-436-966-19	Sequence 19, Appl
23	20	100.0	179 8	US-08-403-590B-77	Sequence 77, Appl
24	20	100.0	179 8	US-08-444-112-77	Sequence 77, Appl
25	20	100.0	254 1	PCT-US00-26395-3	Sequence 3, Appl
26	20	100.0	254 18	US-09-407-430-3	Sequence 3, Appl
27	20	100.0	333 30	US-10-445-724-3	Sequence 3, Appl
28	20	100.0	333 32	US-60-409-909-4	Sequence 4, Appl
29	20	100.0	363 1	PCT-US00-26395-2	Sequence 2, Appl
30	20	100.0	363 18	US-09-407-430-2	Sequence 2, Appl
31	20	100.0	414 3	US-07-748-561-2	Sequence 2, Appl
32	20	100.0	414 3	US-07-989-843-32	Sequence 32, Appl
33	20	100.0	414 8	US-08-463-849-32	Sequence 32, Appl
34	20	100.0	414 8	US-08-463-884-32	Sequence 32, Appl
35	20	100.0	414 20	US-09-690-359-32	Sequence 32, Appl
36	20	100.0	454 8	US-08-403-590B-73	Sequence 73, Appl
37	20	100.0	454 8	US-08-444-112-73	Sequence 73, Appl
38	20	100.0	480 8	US-08-471-498-14	Sequence 14, Appl
39	20	100.0	531 21	US-09-721-480-7	Sequence 7, Appl
40	20	100.0	622 3	US-07-748-561-4	Sequence 4, Appl
41	20	100.0	622 3	US-07-989-843-34	Sequence 34, Appl
42	20	100.0	622 8	US-08-463-849-34	Sequence 34, Appl
43	20	100.0	622 8	US-08-463-884-34	Sequence 34, Appl
-44	20	100.0	622 20	US-09-690-359-34	Sequence 34, Appl
45	20	100.0	637 1	PCT-US02-20676-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-08-974-685-43
; Sequence 43, Application US/08974685

GENERAL INFORMATION:

APPLICANT: LEROUX-ROELS, GEERT
DELEYS, ROBERT
MAERTENS, GEERT

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF

HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 181

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,685
FILING DATE: 19-NOV-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 43:

US-08-974-685-43

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYDYPYR 20
|||||
DB 1 SRCGSGPWITPRCLVDYDYPYR 20

RESULT 2

US-08-974-690-43
; Sequence 43, Application US/08974690
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; APPLICANT: DELEYS, ROBERT
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; TITLE OF INVENTION: HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
; TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 166
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,690
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/635,886
; FILING DATE: 25-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

US-08-974-690-43

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYDYPYR 20
|||||
DB 1 SRCGSGPWITPRCLVDYDYPYR 20

RESULT 3

US-08-974-690A-43
; Sequence 43, Application US/08974690A
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; APPLICANT: DELEYS, ROBERT
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; TITLE OF INVENTION: HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
; TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,690A
; FILING DATE: 19-NOV-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B. J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 2752-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 43:

US-08-974-690A-43

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYDYPYR 20
|||||
DB 1 SRCGSGPWITPRCLVDYDYPYR 20

RESULT 4

US-08-974-690B-43
; Sequence 43, Application US/08974690B
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, GEERT
; APPLICANT: DELEYS, ROBERT
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
; TITLE OF INVENTION: HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND


```

; THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
;
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,690B
; FILING DATE: 19-Nov-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B. J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 2752-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
;
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 43:
US-08-974-690B-43      100.0%; Score 20; DB 13; Length 20;
Query Match           100.0%; Pred. No. 1.2e-13;
Best Local Similarity 100.0%; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

QY      1 SRCGSGPWITPRCLVDYPYR 20
      |||||
DB      1 SRCGSGPWITPRCLVDYPYR 20
      |||||

RESULT 5
US-08-974-690C-43
; Sequence 43, Application US/08974690C
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/08/974,690C
; CURRENT FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-974-690C-43

Query Match           100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SRCGSGPWITPRCLVDYPYR 20
      |||||
DB      1 SRCGSGPWITPRCLVDYPYR 20
      |||||

THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
;
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,690B
; FILING DATE: 19-Nov-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B. J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 2752-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
;
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 43:
US-08-974-690B-43      100.0%; Score 20; DB 13; Length 20;
Query Match           100.0%; Pred. No. 1.2e-13;
Best Local Similarity 100.0%; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

QY      1 SRCGSGPWITPRCLVDYPYR 20
      |||||
DB      1 SRCGSGPWITPRCLVDYPYR 20
      |||||

RESULT 6
US-09-899-303-87
; Sequence 87, Application US/09899303
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: BUYSE, MARIE-ANGE
; APPLICANT: DE MARTYNOFF, GUY
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 87:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 87:
US-09-899-303-87      100.0%; Score 20; DB 23; Length 20;
Query Match           100.0%; Pred. No. 1.2e-13;
Best Local Similarity 100.0%; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

QY      1 SRCGSGPWITPRCLVDYPYR 20
      |||||
DB      1 SRCGSGPWITPRCLVDYPYR 20
      |||||

RESULT 7
US-09-899-303A-87
; Sequence 87, Application US/09899303A
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: BUYSE, MARIE-ANGE
; APPLICANT: DE MARTYNOFF, GUY
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
```

STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (BPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303A
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 87:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 87:
US-09-899-303A-87

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
|||||
DB 1 SRCGSGPWITPRCLVDYPYR 20
|||||

RESULT 8
US-09-973-025-87
; Sequence 87, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 87:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 87:
US-09-973-025-87

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
|||||
DB 1 SRCGSGPWITPRCLVDYPYR 20
|||||

RESULT 9
US-09-995-791-87
; Sequence 87, Application US/09995791
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent in 3.1
; SEQ ID NO 87
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-87

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
|||||
DB 1 SRCGSGPWITPRCLVDYPYR 20
|||||

RESULT 10
US-09-995-808-87
; Sequence 87, Application US/09995808
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent in 3.1
; SEQ ID NO 87
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-87

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;

```
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SRCGSGPWITPRCLVDYPYR 20
   |||||
Db 1 SRCGSGPWITPRCLVDYPYR 20
   |||||

RESULT 11
US-09-995-860-87
; Sequence 87, Application US/0995860
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 87
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-87

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SRCGSGPWITPRCLVDYPYR 20
   |||||
Db 1 SRCGSGPWITPRCLVDYPYR 20
   |||||

RESULT 12
US-10-020-510-87
; Sequence 87, Application US/10020510
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-72
; CURRENT FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 87
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-020-510-87

Query Match 100.0%; Score 20; DB 26; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SRCGSGPWITPRCLVDYPYR 20
   |||||
Db 1 SRCGSGPWITPRCLVDYPYR 20
   |||||

RESULT 13
US-10-321-798-87
; Sequence 87, Application US/10321798
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-93
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: 60/418,358
; PRIOR FILING DATE: 2002-10-16
```

```
; PRIOR APPLICATION NUMBER: 10/020,510
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 87
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-321-798-87

Query Match 100.0%; Score 20; DB 29; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SRCGSGPWITPRCLVDYPYR 20
   |||||
Db 1 SRCGSGPWITPRCLVDYPYR 20
   |||||

RESULT 14
US-08-475-482-44
; Sequence 44, Application US/08475482
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,482
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 17-June-1994
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-475-482-44
```

Job time : 163.1 secs

Query Match 100.0%; Score 20; DB 8; Length 42;
Best Local Similarity 100.0%; Pred. No. 2.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
Db 17 SRCGSGPWITPRCLVDYPYR 36

RESULT 15
US-08-477-072-44
; Sequence 44, Application US/08477072
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,072
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 17-June-1994
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 28-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-477-072-44

Query Match 100.0%; Score 20; DB 8; Length 42;
Best Local Similarity 100.0%; Pred. No. 2.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRCGSGPWITPRCLVDYPYR 20
Db 17 SRCGSGPWITPRCLVDYPYR 36

GenCore version 5.1.6

Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.55 Seconds
(without alignments)
117.536 Million cell updates/sec

Title: US-09-973-025-87

Perfect score: 20

Sequence: 1 SRCGSGPWITPRCLVDYPR 20

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents AA New.*

- 1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	6	US-10-651-165-43
2	20	100.0	68	6	US-10-651-165-250
3	20	100.0	68	6	US-10-651-165-252
4	20	100.0	333	1	PCT-US03-19834-3
5	20	100.0	333	6	US-10-655-562-4
6	20	100.0	637	1	PCT-US03-33610-4
7	20	100.0	3011	1	PCT-US03-19834-2
8	20	100.0	3011	6	US-10-296-734-406
9	19	95.0	30	6	US-10-296-734-486
10	19	95.0	1997	6	US-10-296-734-816
11	19	95.0	5985	6	US-10-296-734-810
12	16	80.0	30	6	US-10-296-734-488
13	11	55.0	68	6	US-10-651-165-254
14	11	55.0	68	6	US-10-651-165-255
15	11	55.0	68	6	US-10-651-165-256
16	11	55.0	68	6	US-10-651-165-262
17	11	55.0	70	6	US-10-651-165-264
18	8	40.0	20	6	US-10-651-165-42
19	8	40.0	20	6	US-10-651-165-44
20	7	35.0	33	6	US-10-685-435-28
21	7	35.0	45	6	US-10-685-435-27
22	7	35.0	68	6	US-10-651-165-251
23	6	30.0	68	6	US-10-651-165-253
24	6	30.0	68	6	US-10-651-165-257
25	6	30.0	68	6	US-10-651-165-258
26	6	30.0	68	6	US-10-651-165-259

27 6 30.0 68 6 US-10-651-165-260 Sequence 260, App
28 6 30.0 68 6 US-10-651-165-261 Sequence 261, App
29 70 6 US-10-651-165-263 Sequence 263, App
30 6 30.0 319 6 US-10-679-063-13689 Sequence 13689, A
31 6 30.0 347 6 US-10-664-391-9 Sequence 9, Appli
32 6 30.0 446 6 US-10-679-063-9258 Sequence 9258, Ap
33 6 30.0 539 6 US-10-664-391-11 Sequence 11, Appli
34 6 30.0 657 6 US-10-425-114A-64717 Sequence 64717, A
35 6 30.0 945 6 US-10-258-899A-1899 Sequence 1899, Ap
36 6 30.0 1015 6 US-10-258-899A-1898 Sequence 1898, Ap
37 6 30.0 1026 1 PCT-US03-20409-3 Sequence 3, Appli
38 6 30.0 1058 6 US-10-258-899A-3866 Sequence 3866, Ap
39 6 30.0 1058 6 US-10-258-899A-3867 Sequence 3867, Ap
40 6 30.0 2280 1 PCT-US03-20322-211 Sequence 211, App
41 6 30.0 3033 6 US-10-009-002-5 Sequence 5, Appli
42 6 30.0 19662 6 US-10-084-846A-6 Sequence 6, Appli
43 6 30.0 19725 6 US-10-084-846A-4 Sequence 4, Appli
44 5 25.0 8 6 US-10-651-165-122 Sequence 122, App
45 5 25.0 8 6 US-10-651-165-123 Sequence 123, App

ALIGNMENTS

RESULT 1

US-10-651-165-43
; Sequence 43, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651.165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974.690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-43

Query Match 100.0%; Score 20; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.4e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPR 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 SRCGSGPWITPRCLVDYPR 20

RESULT 2

US-10-651-165-250
; Sequence 250, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651.165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974.690C
; PRIOR FILING DATE: 1997-11-19

; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 250
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-250

Query Match 100.0%; Score 20; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 2.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
|||||
DB 25 SRCGSGPWITPRCLVDYPYR 44
|||||

RESULT 3
US-10-651-165-252
; Sequence 252, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 252
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-252

Query Match 100.0%; Score 20; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 2.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
|||||
DB 25 SRCGSGPWITPRCLVDYPYR 44
|||||

RESULT 4
PCT-US03-19834-3
; Sequence 3, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; TITLE OF INVENTION: METHODS FOR TREATING HYPERLIPIDEMIA
; FILE REFERENCE: IOWA.045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-3

Query Match 100.0%; Score 20; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 5.9e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
|||||
DB 213 SRCGSGPWITPRCLVDYPYR 232
|||||

RESULT 5
US-10-655-562-4
; Sequence 4, Application US/10655562
; GENERAL INFORMATION:
; APPLICANT: BISHOP-HURLEY, SHARON L.
; APPLICANT: SCHMIDT, FRANCIS J.
; APPLICANT: SMITH, ARNOLD L.
; TITLE OF INVENTION: PHAGE-DISPLAY PEPTIDES AS NOVEL ANTIMICROBIAL AGENTS
; TITLE OF INVENTION: AGAINST HAEMOPHILUS INFLUENZAE
; FILE REFERENCE: UWMO-022US
; CURRENT APPLICATION NUMBER: US/10/655,562
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: 60,409,909
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-10-655-562-4

Query Match 100.0%; Score 20; DB 6; Length 333;
Best Local Similarity 100.0%; Pred. No. 5.9e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
|||||
DB 213 SRCGSGPWITPRCLVDYPYR 232
|||||

RESULT 6
PCT-US03-33610-4
; Sequence 4, Application PC/TUS0333610
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: SELEY, Mark
; APPLICANT: FALIARD, Xavier
; TITLE OF INVENTION: ACTIVATION OF HCV-SPECIFIC T-CELLS
; FILE REFERENCE: 2300-1612.60
; CURRENT APPLICATION NUMBER: PCT/US03/33610
; CURRENT FILING DATE: 2003-10-24
; PRIOR APPLICATION NUMBER: 10/281,341
; PRIOR FILING DATE: 2002-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
PCT-US03-33610-4

```
Query Match          100.0%; Score 20; DB 1; Length 637;
Best Local Similarity 100.0%; Pred. No. 9.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SRCGSGPWITPRCLVDYPYR 20
DB      423 SRCGSGPWITPRCLVDYPYR 442

RESULT 7
PCT-US03-19834-2
; Sequence 2, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; FILE REFERENCE: IOWA:045W0
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-2

Query Match          100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SRCGSGPWITPRCLVDYPYR 20
DB      595 SRCGSGPWITPRCLVDYPYR 614

RESULT 8
US-10-296-734-406
; Sequence 406, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la consensus polyprotein
US-10-296-734-406

Query Match          100.0%; Score 20; DB 6; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SRCGSGPWITPRCLVDYPYR 20
DB      595 SRCGSGPWITPRCLVDYPYR 614

RESULT 9
US-10-296-734-486
; Sequence 486, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 486
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la segment 40
US-10-296-734-486

Query Match          95.0%; Score 19; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.2e-12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SRCGSGPWITPRCLVDYPY 19
DB      12 SRCGSGPWITPRCLVDYPY 30

RESULT 10
US-10-296-734-816
; Sequence 816, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 816
; LENGTH: 1997
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette C
US-10-296-734-816

Query Match          95.0%; Score 19; DB 6; Length 1997;
Best Local Similarity 100.0%; Pred. No. 1.9e-11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SRCGSGPWITPRCLVDYPY 19
DB      1842 SRCGSGPWITPRCLVDYPY 1860

RESULT 11
US-10-296-734-810
; Sequence 810, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
```

```
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 810
; LENGTH: 5985
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la savine
US-10-296-734-810

Query Match          95.0%; Score 19; DB 6; Length 5985;
Best Local Similarity 100.0%; Pred. No. 3.8e-11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRGSGPWITPRCLVDYPY 19
   |||||
Db 5832 SRGSGPWITPRCLVDYPY 5850

RESULT 12
US-10-296-734-488
; Sequence 488, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 488
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la segment 41
US-10-296-734-488

Query Match          80.0%; Score 16; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SGFWITPRCLVDYPYR 20
   |||||
Db 1 SGFWITPRCLVDYPYR 16

RESULT 13
US-10-651-165-254
; Sequence 254, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 254
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-256

; ORGANISM: hepatitis C virus
US-10-651-165-254

Query Match          55.0%; Score 11; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 TPRCLVDYPYR 20
   |||||
Db 34 TPRCLVDYPYR 44

RESULT 14
US-10-651-165-255
; Sequence 255, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 255
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-255

Query Match          55.0%; Score 11; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 TPRCLVDYPYR 20
   |||||
Db 34 TPRCLVDYPYR 44

RESULT 15
US-10-651-165-256
; Sequence 256, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 256
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-256
```


Query Match 55.0%; Score 11; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 TPRCLVDYDYPYR 20
|||
Db 34 TPRCLVDYDYPYR 44
|||

Search completed: November 21, 2003, 22:12:58
Job time : 8.55 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:46:27 ; Search time 9.25 Seconds
(without alignments)
207.932 Million cell updates/sec

Title: US-09-973-025-87

Perfect score: 20

Sequence: 1 SRCGSGPWITPRCLVDYPYR 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

1: PIR 76:*

2: pir1:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	640	2 JQ1584	genome polyprotein
2	20	100.0	3011	1 GNVWC3	genome polyprotein
3	20	100.0	3011	1 S40770	genome polyprotein
4	15	75.0	716	2 JQ1366	polyprotein - hepa
5	11	55.0	350	2 S35631	genome polyprotein
6	11	55.0	3033	1 GNVWJ8	hypothetical prote
7	7	35.0	371	2 T30762	genome polyprotein
8	7	35.0	3011	1 GNVVCH	genome polyprotein
9	7	35.0	3014	1 JCS620	genome polyprotein
10	6	30.0	114	1 NRUSU2	ribonuclease U2 (
11	6	30.0	234	2 S32742	genome polyprotein
12	6	30.0	235	2 S32747	genome polyprotein
13	6	30.0	284	2 AH3570	nickel transport s
14	6	30.0	295	1 H70031	conserved hypothet
15	6	30.0	782	2 S19876	genome polyprotein
16	6	30.0	782	2 S18031	genome polyprotein
17	6	30.0	782	2 S18032	genome polyprotein
18	6	30.0	787	2 PNO677	genome polyprotein
19	6	30.0	782	2 S19875	hypothetical prote
20	6	30.0	875	2 T30023	hypothetical prote
21	6	30.0	935	2 T23390	hypothetical prote
22	6	30.0	1015	2 S53474	Human giant larvae
23	6	30.0	1153	2 T21386	hypothetical prote
24	6	30.0	1562	2 T23146	hypothetical prote
25	6	30.0	3010	1 GNVVTC	genome polyprotein
26	6	30.0	3010	1 GNVWVC	genome polyprotein
27	6	30.0	3010	1 A45573	genome polyprotein
28	6	30.0	3010	1 S18030	genome polyprotein
29	6	30.0	3010	1 GNVWTV	genome polyprotein

30	6	30.0	3033	1 JQ1303	genome polyprotein
31	5	25.0	31	2 S33176	microtubule-associ
32	5	25.0	48	2 E48110	RNA recognition mo
33	5	25.0	68	2 A95993	hypothetical prote
34	5	25.0	88	2 S36917	ubiquinol-cytochro
35	5	25.0	91	2 T00138	hypothetical prote
36	5	25.0	112	2 D72748	hypothetical prote
37	5	25.0	119	2 T49363	hypothetical prote
38	5	25.0	122	2 G70970	hypothetical prote
39	5	25.0	125	2 I52650	microtubule-associ
40	5	25.0	126	2 A69250	hypothetical prote
41	5	25.0	131	2 A99913	hypothetical prote
42	5	25.0	133	2 D83123	hypothetical prote
43	5	25.0	133	2 T46260	hypothetical prote
44	5	25.0	135	2 S78394	ribosomal protein
45	5	25.0	137	2 T12880	hypothetical prote

ALIGNMENTS

RESULT 1

JQ1584

genome polyprotein - hepatitis C virus (strain U.K.) (fragment)

N:Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural prot

C:Species: hepatitis C virus

C>Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000

C:Accession: JQ1584

R:Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.

J. Gen. Virol. 73, 1521-1525, 1992

A:Title: Cloning and sequencing of the structural region and expression of putative core

A:Reference number: JQ1584; MUID:92300349; PMID:1318944

A:Accession: JQ1584

A:Molecule type: genomic RNA

A:Residues: 1-640 <KUM>

A:Cross-references: GB:X84079; NID:G643119; PIDN:CAA58888.1; PID:G643120

A:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; polype

F:1-191/Product: core protein C #status predicted <CPC>

F:192-389/Product: envelope protein E1 #status predicted <EE1>

F:390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predicted <

F:196,209,234,305,417,430,448,476,540,556,576,632/Binding site: carbohydrate (Asn) (cov

Query Match 100.0%; Score 20; DB 2; Length 640;

Best Local Similarity 100.0%; Pred. No. 7.3e-15;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20

DB 595 SRCGSGPWITPRCLVDYPYR 614

RESULT 2

GNWVC3

genome polyprotein - hepatitis C virus (strain HCV-1)

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructu

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001

C:Accession: A39166; PQ0403; PQ0404

R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi

Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991

A:Title: Genetic organization and diversity of the hepatitis C virus.

A:Reference number: A39166; MUID:91172826; PMID:1848704

A:Accession: A39166

A:Molecule type: mRNA

A:Residues: 1-3011 <CHO>

A:Cross-references: GB:M62321; NID:G329873; PIDN:AAA45676.1; PID:G329874

R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.J

J. Gen. Virol. 73, 1131-1141, 1992

A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to

A:Reference number: PQ0393; MUID:92268871; PMID:1316939

A:Accession: PQ0403

A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CHA>
A:Cross-references: DDBJ:D10128
A:Experimental source: isolates E-b16
A:Accession: PQ0404
A>Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CH2>
A:Experimental source: isolates E-b17
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F:115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis virus #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196-209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22
Query Match 100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
|||||
Db 595 SRCGSGPWITPRCLVDYPYR 614

RESULT 3
S40770
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: S40770; PC1285
R:Okamoto, H.
submitted to the EMBL Data Library, March 1992
A:Reference number: S40770
A:Accession: S40770
A:Molecule type: genomic RNA
A:Residues: 1-3011 <OKA>
A:Cross-references: EMBL:D10749; NID:G221586; PIDN:BAA01582.1; PID:G221587
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; MUID:91013116; PMID:2170712
A:Accession: PC1285
A:Molecule type: genomic RNA
A:Residues: 1-513 <OK2>
A:Cross-references: GB:D00831; NID:G221511; PIDN:BAA00705.1; PID:G221512
A:Experimental source: isolate HC-J1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis virus #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
|||||
Db 595 SRCGSGPWITPRCLVDYPYR 614

RESULT 4
JQ1366
polyprotein - hepatitis C virus (French isolate) (fragments)
C:Species: hepatitis C virus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C:Accession: JQ1366
R:Kremsdorf, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.
J. Gen. Virol. 72, 2557-2561, 1991
A:Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication
A:Reference number: JQ1366; MUID:92013977; PMID:1655961
A:Accession: JQ1366
A:Molecule type: genomic RNA
A:Residues: 1-716 <KRE>
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: glycoprotein; polyprotein
F:84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #sta
Query Match 75.0%; Score 15; DB 2; Length 716;
Best Local Similarity 100.0%; Pred. No. 3.3e-09;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLV 15
|||||
Db 262 SRCGSGPWITPRCLV 276

RESULT 5
S35631
genome polyprotein - hepatitis C virus (fragment)
N:Contains: nonstructural protein 1
C:Species: hepatitis C virus
C:Date: 19-Mar-1997 #sequence_revision 30-Jan-1998 #text_change 17-Nov-2000
C:Accession: S35631
R:Sarashina, T.; Sakurai, T.; Watanabe, Y.; Kashima, K.; Suzuki, T.; Chiba, J.; Kita, Y.
Nucleic Acids Res. 21, 1037, 1993
A:Title: Nucleotide sequence of the hepatitis C virus genome from a patient negative for
A:Reference number: S35629; MUID:93197128; PMID:8383835
A:Accession: S35631
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-350 <SAR>
A:Cross-references: EMBL:D13970
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1992
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein
F:1-350/Product: nonstructural protein 1 #status predicted <MAT>
Query Match 55.0%; Score 11; DB 2; Length 350;
Best Local Similarity 100.0%; Pred. No. 5.5e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 TPRCLVDYPYR 20
|||||
Db 225 TPRCLVDYPYR 235

RESULT 6
GNVJ98
genome polyprotein - hepatitis C virus (strain HC-J8)
N:Contains: capsid protein C; envelope protein M; hepatitis virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40250; PQ0397; PQ0559

QY 10 TPRLCLVD 16
| | | | |
DB 126 TPRLCLVD 132

RESULT 8
GNMWCH
genome polypeptide - hepatitis C virus (strain H)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
R;Accession: A36814; A41546
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A;Description: Genomic structure of the human prototype strain H of hepatitis C virus: comparison with other human and chimpanzee strains
A;Reference number: A36814
A;Accession: A36814
A;Molecule type: genomic RNA
A;Residues: 1-3011 <NC>
A;Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: comparison with other human and chimpanzee strains
A;Reference number: A41546; MUID:92052256; PMID:1658800
A;Contents: annotation
A;Note: neither amino acid nor nucleotide sequence is given
C;Superfamily: hepatitis C virus genome polypeptide
C;Keywords: AP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,2303,2304,2305,2306,2307,2308,2309,2310,2311,2312,2313,2314,2315,2316,2317,2318,2319,2320,2321,2322,2323,2324,2325,2326,2327,2328,2329,2330,2331,2332,2333,2334,2335,2336,2337,2338,2339,2340,2341,2342,2343,2344,2345,2346,2347,2348,2349,2350,2351,2352,2353,2354,2355,2356,2357,2358,2359,2360,2361,2362,2363,2364,2365,2366,2367,2368,2369,2370,2371,2372,2373,2374,2375,2376,2377,2378,2379,2380,2381,2382,2383,2384,2385,2386,2387,2388,2389,2390,2391,2392,2393,2394,2395,2396,2397,2398,2399,2400,2401,2402,2403,2404,2405,2406,2407,2408,2409,2410,2411,2412,2413,2414,2415,2416,2417,2418,2419,2420,2421,2422,2423,2424,2425,2426,2427,2428,2429,2430,2431,2432,2433,2434,2435,2436,2437,2438,2439,2440,2441,2442,2443,2444,2445,2446,2447,2448,2449,2450,2451,2452,2453,2454,2455,2456,2457,2458,2459,2460,2461,2462,2463,2464,2465,2466,2467,2468,2469,2470,2471,2472,2473,2474,2475,2476,2477,2478,2479,2480,2481,2482,2483,2484,2485,2486,2487,2488,2489,2490,2491,2492,2493,2494,2495,2496,2497,2498,2499,2500,2501,2502,2503,2504,2505,2506,2507,2508,2509,2510,2511,2512,2513,2514,2515,2516,2517,2518,2519,2520,2521,2522,2523,2524,2525,2526,2527,2528,2529,2530,2531,2532,2533,2534,2535,2536,2537,2538,2539,2540,2541,2542,2543,2544,2545,2546,2547,2548,2549,2550,2551,2552,2553,2554,2555,2556,2557,2558,2559,2560,2561,2562,2563,2564,2565,2566,2567,2568,2569,2570,2571,2572,2573,2574,2575,2576,2577,2578,2579,2580,2581,2582,2583,2584,2585,2586,2587,2588,2589,2590,2591,2592,2593,2594,2595,2596,2597,2598,2599,2600,2601,2602,2603,2604,2605,2606,2607,2608,2609,2610,2611,2612,2613,2614,2615,2616,2617,2618,2619,2620,2621,2622,2623,2624,2625,2626,2627,2628,2629,2630,2631,2632,2633,2634,2635,2636,2637,2638,2639,2640,2641,2642,2643,2644,2645,2646,2647,2648,2649,2650,2651,2652,2653,2654,2655,2656,2657,2658,2659,2660,2661,2662,2663,2664,2665,2666,2667,2668,2669,2670,2671,2672,2673,2674,2675,2676,2677,2678,2679,2680,2681,2682,2683,2684,2685,2686,2687,2688,2689,2690,2691,2692,2693,2694,2695,2696,2697,2698,2699,2700,2701,2702,2703,2704,2705,2706,2707,2708,2709,2710,2711,2712,2713,2714,2715,2716,2717,2718,2719,2720,2721,2722,2723,2724,2725,2726,2727,2728,2729,2730,2731,2732,2733,2734,2735,2736,2737,2738,2739,2740,2741,2742,2743,2744,2745,2746,2747,2748,2749,2750,2751,2752,2753,2754,2755,2756,2757,2758,2759,2760,2761,2762,2763,2764,2765,2766,2767,2768,2769,2770,2771,2772,2773,2774,2775,2776,2777,2778,2779,2780,2781,2782,2783,2784,2785,2786,2787,2788,2789,2790,2791,2792,2793,2794,2795,2796,2797,2798,2799,2800,2801,2802,2803,2804,2805,2806,2807,2808,2809,2810,2811,2812,2813,2814,2815,2816,2817,2818,2819,2820,2821,2822,2823,2824,2825,2826,2827,2828,2829,2830,2831,2832,2833,2834,2835,2836,2837,2838,2839,2840,2841,2842,2843,2844,2845,2846,2847,2848,2849,2850,2851,2852,2853,2854,2855,2856,2857,2858,2859,2860,2861,2862,2863,2864,2865,2866,2867,2868,2869,2870,2871,2872,2873,2874,2875,2876,2877,2878,2879,2880,2881,2882,2883,2884,2885,2886,2887,2888,2889,2890,2891,2892,2893,2894,2895,2896,2897,2898,2899,2900,2901,2902,2903,2904,2905,2906,2907,2908,2909,2910,2911,2912,2913,2914,2915,2916,2917,2918,2919,2920,2921,2922,292

F:12-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEP>
F:384-408/Region: hypervariable #status predicted
F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>
F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
F:1008-1616/Product: hepatitis virus #status predicted <NS3>
F:1231-1238/Region: nucleotide-binding motif A (P-loop)
F:1313-1318/Region: nucleotide-binding motif B
F:1317-1320/Region: DEXH motif
F:1617-1863/Product: nonstructural protein NS4a #status predicted <N4A>
F:1864-2014/Product: nonstructural protein NS4b #status predicted <N4B>
F:2013-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F:2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 35.0%; Score 7; DB 1; Length 3014;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 RCGSGPW 8
|||||
Db 597 RCGSGPW 603

RESULT 10
NRUSU2
N:ribonuclease U2 (EC 3.1.1.27.4) [validated] - smut fungus (Ustilago sphaerogena)
N:Alternate names: ribonuclease U2 isoform A; ribonuclease U2 isoform B
C:Species: Ustilago sphaerogena
C:Date: 24-Apr-1984 #sequence revision 17-Sep-1997 #text_change 03-Jun-2002
C:Accession: PC4081; A26071; A00801
R:Kanaya, S.; Uchida, T.
J. Biochem. 118, 681-682, 1995
A:Title: Revised sequence of ribonuclease U2 in the substrate-binding region.
A:Reference number: PC4081; MUID:96157700; PMID:8576077

A:Accession: PC4081
A:Molecule type: protein
A:Residues: 38-65 <KAN>
R:Kanaya, S.; Uchida, T.
Biochem. J. 240, 163-170, 1986
A:Title: Comparison of primary structures of ribonuclease U2 isoforms.
A:Reference number: A26071; MUID:87156566; PMID:3827836
A:Accession: A26071
A:Molecule type: protein
A:Residues: 1-48, 'DQ', 51-114 <KA2>
R:Sato, S.; Uchida, T.
Biochem. J. 145, 353-360, 1975

A:Title: The amino acid sequence of ribonuclease U-2 from Ustilago sphaerogena.
A:Reference number: A90280; MUID:75224613; PMID:1156364
A:Accession: A00801
A:Molecule type: protein
A:Residues: 1, 'N', 3-4, 'E', 6-31, 'RP', 34, 'G', 37-48, 'DQ', 51-56, 'P', 58, 'S', 60-114 <SAT>
R:Sato, S.; Uchida, T.
J. Biochem. 77, 1171-1176, 1975
A:Title: The disulfide bridges of ribonuclease U-2 from Ustilago sphaerogena.
A:Reference number: A91936; MUID:76189971; PMID:1225902
A:Contents: annotation; disulfide bonds
R:Uchida, T.; Sato, S.
in Ribosomes and RNA Metabolism, Zelinka, J., and Balan, J., eds., pp. 453-472, Publ. Hou

A:Reference number: A9440
A:Contents: annotation; active site
R:Noguchi, S.; Satow, Y.; Uchida, T.; Sasaki, C.; Matsuzaki, T.
submitted to the Brookhaven Protein Data Bank, May 1995
A:Reference number: A66564; PDB:1RTU
A:Contents: annotation; X-ray crystallography, 1.8 angstroms
C:Comment: This enzyme, a purine-specific ribonuclease, exists in two isoforms, U2-A and U2-B
C:Comment: This fungus is a basidiomycete.
C:Comment: Isoform B, a less active form, results from the isopeptide isomerization of the
C:Superfamily: ribonuclease T1
C:Keywords: endonuclease; extracellular protein; hydrolase
F:1-54, 9-113-55-96/Disulfide bonds: #status experimental
F:39, 41, 85, 101/Active site: Tyr, His, Arg, His #status predicted
F:62/Active site: Glu #status experimental

Query Match 30.0%; Score 6; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CGSGPW 8
|||||
Db 55 CGSGPW 60

RESULT 11
S32742
genome polyprotein - hepatitis C virus (isolate CR-1) (fragment)
N:Contains: envelope protein E2
C:Species: hepatitis C virus
A:Variety: isolate CR-1
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S32742
R:Roggenbier, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
submitted to the EMBL Data Library, February 1993
A:Description: Variability of the envelope regions of HCV in European isolates and its
A:Reference number: S32741
A:Accession: S32742
A:Molecule type: genomic RNA
A:Residues: 1-234 <ROG>
A:Cross-references: EMBL:X72979; NID:g296102; PIDN:CAA51484.1; PID:g296103
A:Experimental source: isolate CR-1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; polyprotein
F:1-234/Product: envelope protein E2 #status predicted <MAT>

Query Match 30.0%; Score 6; DB 2; Length 234;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 CGSGPW 8
|||||
Db 228 CGSGPW 233

RESULT 12
S32747
genome polyprotein - hepatitis C virus (isolate HU-1) (fragment)
N:Contains: envelope protein E2
C:Species: hepatitis C virus
A:Variety: isolate HU-1
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S32747
R:Roggenbier, M.; Lu, M.; Fuchs, G.; Ernst, G.; Hoehne, M.; Schreier, E.
submitted to the EMBL Data Library, February 1993
A:Description: Variability of the envelope regions of HCV in European isolates and its
A:Reference number: S32741
A:Accession: S32747
A:Molecule type: genomic RNA
A:Residues: 1-235 <ROG>
A:Cross-references: EMBL:X72977; NID:g296112; PIDN:CAA51482.1; PID:g296113
A:Experimental source: isolate HU-1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: envelope protein; polyprotein
F:1-235/Product: envelope protein E2 #status predicted <MAT>

Query Match 30.0%; Score 6; DB 2; Length 235;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CGSGPW 8
|||||
Db 229 CGSGPW 234

RESULT 13
AH3570
nickel transport system permease protein nikC [imported] - Brucella melitensis (strain

C:Species: Brucella melitensis
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
 C:Accession: AH3570
 R:DeVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, E.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
 A:Reference number: AD3252; PMID:11756688
 A:Accession: AH3570
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-284 <KUR>
 A:Cross-references: GB:AE008918; PIDN:AAL53731.1; PID:gl7984656; GSPDB:GN00191
 A:Experimental source: strain 16M
 C:Genetics:
 A:Gene: BMEI10489
 A:Map position: II
 C:Superfamily: oligopeptide permease protein oppB

Query Match 30.0%; Score 6; DB 2; Length 284;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GPWITP 11
 |||||
 DB 35 GPWITP 40

RESULT 14
 H70031
 conserved hypothetical protein yvcJ - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C:Accession: H70031
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
 Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seroz
 akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033; PMID:9384377
 A:Accession: H70031
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-295 <KUN>
 A:Cross-references: GB:Z99121; GB:AL009126; NID:g2635827; PIDN:CAB15482.1; PID:g26355990
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yvcJ
 C:Superfamily: Bacillus subtilis conserved hypothetical protein yvcJ

Query Match 30.0%; Score 6; DB 1; Length 295;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 PWITPR 12
 |||||
 DB 88 PWITPR 93

RESULT 15
 S19876
 genome polypeptide - hepatitis C virus (isolate JKS) (fragment)
 N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein

C:Species: hepatitis C virus
 A:Variety: isolate JKS
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
 C:Accession: S19876
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
 A:Reference number: S18029
 A:Accession: S19876
 A:Molecule type: genomic RNA
 A:Residues: 1-782 <HON>
 A:Cross-references: EMBL:X61595; NID:g59486; PIDN:CAA43792.1; PID:g59487
 A:Experimental source: isolate JKS
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
 F:1-191/Product: core protein #status predicted <MAT1>
 F:192-383/Product: envelope protein 1 #status predicted <MAT2>
 F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
 F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 30.0%; Score 6; DB 2; Length 782;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CGSGPW 8
 |||||
 DB 597 CGSGPW 602

Search completed: November 21, 2003, 21:11:37
 Job time : 10.25 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:52 ; Search time 4.85 Seconds
(without alignments)
193.925 Million cell updates/sec

Title: US-09-973-025-87

Perfect score: 20

Sequence: 1 SRCGSPWITPRLVDVYPYR 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	3011	1	POLG_HCV1
2	11	55.0	737	1	POLG_HCVJ7
3	11	55.0	3033	1	POLG_HCVJ8
4	7	35.0	3011	1	POLG_HCVH
5	6	30.0	114	1	RNU2_USTSP
6	6	30.0	295	1	YVCJ_BACSU
7	6	30.0	601	1	FL3C_MOUSE
8	6	30.0	737	1	POLG_HCVJ5
9	6	30.0	3010	1	POLG_HCVBK
10	6	30.0	3010	1	POLG_HCVJA
11	6	30.0	3010	1	POLG_HCVJT
12	6	30.0	3010	1	POLG_HCVTW
13	6	30.0	3033	1	POLG_HCVJ6
14	5	25.0	89	1	CYB_BRANA
15	5	25.0	126	1	Y001_ARCFU
16	5	25.0	135	1	RK16_EPIVI
17	5	25.0	149	1	GTH2_CLUPA
18	5	25.0	151	1	VG13_HSVSA
19	5	25.0	151	1	VG13_HSVSC
20	5	25.0	157	1	YAL9_PSEAE
21	5	25.0	208	1	Y0Y9_CAEEL
22	5	25.0	209	1	HAMI_CHLMU
23	5	25.0	209	1	HAMI_CHLTR
24	5	25.0	211	1	HPRT_LEIDO
25	5	25.0	215	1	US11_HCNVA
26	5	25.0	216	1	CSGD_ECOLI
27	5	25.0	216	1	CSGD_SALTY
28	5	25.0	218	1	VGLL_HSVBE
29	5	25.0	224	1	NO51_SOYBN
30	5	25.0	253	1	HRB3_XANCV
31	5	25.0	254	1	IBF4_MOUSE
32	5	25.0	254	1	IBP4_RAT
33	5	25.0	258	1	IBP4_BOVIN

RESULT 1

ID	POLG_HCV1	STANDARD;	PRT;	3011 AA.
AC	P26664;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].			
DE	Hepatitis C virus (isolate 1) (HCV).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.			
OX	NCBI_TaxID=11104;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91172826; PubMed=1848704;			
RA	Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;			
RA	"Genetic organization and diversity of the hepatitis C virus.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).			
CC	-!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.			
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	-!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA}(N).			
CC	-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.			
CC	-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; M62321; AAA45676.1; -.			
DR	PIR; A39166; GNWVC3.			
DR	PDB; 1A1V; 16-FEB-99.			
DR	PDB; 1HEI; 25-NOV-98.			
DR	MEROPS; S29.001; -.			
DR	MEROPS; U39.001; -.			
DR	InterPro; IPR001410; DEAD.			
DR	InterPro; IPR002522; HCV_capsid.			

FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 737 737
 SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122F2 CRC64;
 Query Match 55.0%; Score 11; DB 1; Length 737;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 10 TPRCLVDYDYPYR 20
 DB 608 TPRCLVDYDYPYR 618
 RESULT 3
 POLG_HCVJ08 STANDARD; PRT; 3033 AA.
 AC P26661;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin).
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J8) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11115;
 RN [1]_TaxID=11115;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto K., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
 RA Fukuda S., Tsuda F., Mishihiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D10988; BAA01761.1; -;
 CC PIR; A40250; GNWJ08.
 CC HSP; P27958; 1HEI.
 CC MEROPS; S29.001; -;
 CC MEROPS; U39.001; -;
 CC InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXBC; 1.
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2359 2359
 FT CARBOHYD 2811 2811
 SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;
 Query Match 55.0%; Score 11; DB 1; Length 3033;
 Best Local Similarity 100.0%; Pred. No. 0.00034;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 10 TPRCLVDYDYPYR 20
 DB 608 TPRCLVDYDYPYR 618

```
RESULT 4
POLG HCVH STANDARD; PRT; 3011 AA.
ID POLG HCVH STANDARD; PRT; 3011 AA.
AC P2795g;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate H) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11108;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658800;
RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C
RT virus: comparison with American and Japanese isolates.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN [2]
RX X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX MEDLINE=97331322; PubMed=9187654;
RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain.";
RL Nat. Struct. Biol. 4:463-467(1997).
RN [3]
RX X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE=98154321; PubMed=9493270;
RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding.";
RL Structure 6:89-100(1998).
CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; M67463; AAA45534.1; --
DR PIR; A36814; GNVVCH.
DR PDB; 1HEI; 25-NOV-98.
DR PDB; 1A1V; 16-FEB-99.
DR PDB; 1A1R; 17-JUN-98.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR TRANSFAC; T04155; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4A.
DR InterPro; IPR001490; HCV NS4B.
DR InterPro; IPR002868; HCV NS5A.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_p.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4A; 1.
DR Pfam; PF01001; HCV NS4B; 1.
DR Pfam; PF01506; HCV NS5A; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
FT CHAIN 192 383 CAPSID PROTEIN C.
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
FT CHAIN 810 1026 PROTEIN P7.
FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5A.
FT CHAIN 3012 369 NONSTRUCTURAL PROTEIN NS5B.
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 136 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 224 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1224 1226 N-LINKED (GLCNAC. .) (POTENTIAL).
FT TURN 1232 1233
```

FT TURN 1236 1238
FT TURN 1239 1246
FT TURN 1247 1248
FT STRAND 1251 1255
FT STRAND 1258 1271
FT TURN 1272 1272
FT STRAND 1272 1280
FT TURN 1281 1282
FT STRAND 1283 1285
FT STRAND 1291 1295
FT STRAND 1296 1301
FT TURN 1302 1303
FT STRAND 1312 1316
FT TURN 1317 1319
FT TURN 1323 1335
FT TURN 1336 1340
FT STRAND 1343 1347
FT TURN 1352 1353
FT TURN 1360 1361
FT STRAND 1362 1366
FT STRAND 1368 1368
FT STRAND 1373 1375
FT TURN 1376 1377
FT STRAND 1378 1380
FT HELIX 1382 1385
FT STRAND 1389 1393
FT TURN 1397 1409
FT TURN 1410 1411
FT STRAND 1414 1417
FT TURN 1419 1420
FT STRAND 1432 1436
FT TURN 1438 1439
FT STRAND 1450 1453
FT STRAND 1456 1463
FT STRAND 1471 1478
FT STRAND 1480 1480
FT HELIX 1481 1488
FT TURN 1489 1490
FT STRAND 1497 1501
FT STRAND 1507 1507
FT STRAND 1511 1511
FT STRAND 1511 1511
FT HELIX 1514 1527
FT STRAND 1532 1544
FT STRAND 1550 1550
FT HELIX 1555 1564
FT STRAND 1570 1578
FT TURN 1579 1580
FT HELIX 1584 1597
FT TURN 1598 1598
FT HELIX 1606 1611
FT TURN 1614 1618
FT STRAND 1622 1623
FT STRAND 1627 1627
FT STRAND 1635 1636
FT HELIX 1640 1652
SQ SEQUENCE 3011 AA; 327142 MW; 772CBB23CCD94753 CRC64;
Query Match 35.0%; Score 7; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SRCGSGP 7
|||||
Db 595 SRCGSGP 601
RESULT 5
RNU2_USTSP
ID RNU2_USTSP STANDARD; PRT; 114 AA.
AC P00654: O9HGK7;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Ribonuclease U2 (EC 3.1.1.27.4) (RNase U2).
GN RNU2.
OS Ustilago sphaerogena (Smut fungus).
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_TaxID=5271;
RN [1]
RP SEQUENCE.
RX MEDLINE=75224613; PubMed=1156364;
RA Sato S., Uchida T.;
RT "The amino acid sequence of ribonuclease U2 from Ustilago
RT sphaerogena.";
RL Biochem. J. 145:353-360(1975).
RN [2]
RP SEQUENCE OF 15-106 FROM N.A.
RC STRAIN=ATCC 12421;
RX MEDLINE=20389598; PubMed=10930732;
RA Martinez-Ruiz A., Garcia-Ortega L., Kao R., Onaderra M.,
RA Mancheno J.M., Davies J., Martinez del Pozo A., Gavilanes J.G.;
RT "Ribonuclease U2: cloning, production in Pichia pastoris and affinity
RT chromatography purification of the active recombinant protein.";
RL FEMS Microbiol. Lett. 189:165-169(2000).
RN [3]
RP SEQUENCE OF 44-53, AND REVISIONS.
RX MEDLINE=96157700; PubMed=8576077;
RA Kanaya S., Uchida T.;
RT "Revised sequence of ribonuclease U2 in the substrate-binding
RT region.";
RL J. Biochem. 118:681-682(1995).
RN [4]
RP DISULFIDE BONDS.
RX MEDLINE=76189971; PubMed=1225902;
RA Sato S., Uchida T.;
RT "The disulfide bridges of ribonuclease U2 from Ustilago sphaerogena.";
RL J. Biochem. 77:1171-1176(1975).
RN [5]
RP ACTIVE SITE.
RA Uchida T., Sato S.;
RL (In) Zelinka J., Balan J. (eds.);
RL Ribosomes and RNA metabolism, pp.453-472, Publ. House Slovak Acad.
RL Sci., Bratislava (1973).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS), AND REVISIONS.
RX MEDLINE=96082150; PubMed=7492561;
RA Noguchi S., Satow Y., Uchida T., Sasaki C., Matsuzaki T.;
RT "Crystal structure of Ustilago sphaerogena ribonuclease U2 at 1.8-A
RT resolution.";
RL Biochemistry 34:15583-15591(1995).
CC -|- CATALYTIC ACTIVITY: Two-stage endonucleolytic cleavage to
CC nucleoside 3'-phosphates and 3'-phosphooligonucleotides ending in
CC A-P or G-P with 2',3'-cyclic phosphate intermediates.
CC -|- MISCELLANEOUS: After treatment by base Asp-32 and Asp-45 partially
CC isomerise by succinimide rearrangement to form isospartyl
CC peptides.
CC -|- SIMILARITY: BELONGS TO THE RIBONUCLEASE U2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ004827; CAC04098.1; -.
CC PIR; PC4081; NRUSU2.
CC PDB; 1RTU; 08-NOV-96.
CC InterPro; IPR000026; Ribonuc N1T1.
CC Pfam; PF00545; ribonuclease; 1.
KW Hydrolase; Nuclease; Endonuclease; 3D-structure.
FT DISULFID 1 54
FT DISULFID 9 113
FT DISULFID 55 96

FT ACT_SITE 41 41 BY SIMILARITY.
 FT ACT_SITE 62 62 METHYLATION INACTIVATES ENZYME.
 FT ACT_SITE 85 85 BY SIMILARITY.
 FT ACT_SITE 101 101 BY SIMILARITY.
 FT STRAND 6 9
 FT TURN 10 11
 FT STRAND 12 15
 FT STRAND 16 31
 FT TURN 32 33
 FT TURN 36 38
 FT STRAND 41 43
 FT STRAND 60 64
 FT TURN 67 68
 FT STRAND 74 74
 FT TURN 75 76
 FT STRAND 77 77
 FT STRAND 84 89
 FT TURN 90 92
 FT STRAND 95 100
 FT TURN 102 103
 FT TURN 107 108
 FT STRAND 111 112
 SQ SEQUENCE 114 AA; 96D32CB2E23AAB98 CRC64;
 Query Match 30.0%; Score 6; DB 1; Length 114;
 Best Local Similarity 100.0%; Pred. No. 5.6;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 CGSGPW 8
 Db |||||
 55 CGSGPW 60

RESULT 6
 YVCJ_BACSU STANDARD; PRT; 295 AA.
 ID YVCJ_BACSU
 AC Q06973;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hypothetical UPF0042 protein yvcJ.
 GN YVCJ
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OC NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Denizot F.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 Brucillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
 Gim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 Guiseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
 Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
 Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
 Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Soldo B.,
 Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 Takeuchi M., Tamakoshi A., Tanaka T., Terptria P., Tognoni A.,
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
 Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 subtilis.";
 RL Nature 390:249-256 (1997).
 CC - SIMILARITY: Belongs to the UPF0042 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC
 DR EMBL; Z94043; CAB08057.1; -;
 DR EMBL; Z99121; CAB15482.1; -;
 DR PIR; H70031; H70031.
 DR Subtilist; BG12400; yvcJ.
 DR HAMAP; MF_00636; -; 1.
 DR InterPro; IPR005337; UPF0042.
 DR Pfam; PF03668; ATP bind2; 1.
 KW Hypothetical protein; ATP-binding; Complete proteome.
 FT NP_BIND 16 23 ATP (POTENTIAL)
 SQ SEQUENCE 295 AA; 33851 MW; 4D71DCB4459F92A CRC64;
 Query Match 30.0%; Score 6; DB 1; Length 295;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 PWITPR 12
 Db |||||
 88 PWITPR 93

RESULT 7
 FL3C_MOUSE STANDARD; PRT; 601 AA.
 ID FL3C_MOUSE
 AC Q9DBR2;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Protein FAM13C1.
 GN FAM13C1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Lung;
 RX MEDLINE=21085650; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
 Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 CC -!- SIMILARITY: Belongs to the FAM13 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AK004797; BAB23571.1; --
 DR MGD; MGI:1918971; 1200015N20Rik.
 SQ SEQUENCE 601 AA; 66980 MW; ABICADCF04322BDD CRC64;
 Query Match 30.0%; Score 6; DB 1; Length 601;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SRCGSG 6
 Db 293 SRCGSG 298
 POLG_HCVJ5
 ID POLG_HCVJ5 STANDARD; PRT; 737 AA.
 AC P27960;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
 DE protein (Envelope protein M); Major envelope protein E; Nonstructural
 DE proteins NS1 and NS2] (Fragment).
 OS Hepatitis C virus (isolate HC-J5) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11112;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Teuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D10075; BAA00969.1; --
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.

DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT NON_TER 737 737
 SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CRC64;
 Query Match 30.0%; Score 6; DB 1; Length 737;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 CGSGPW 8
 Db 601 CGSGPW 606
 POLG_HCVBK
 ID POLG_HCVBK STANDARD; PRT; 3010 AA.
 AC P26563;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate BK) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11105;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91140698; PubMed=1847440;
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 RA Onishi E., Andoh T., Yoshida I., Okayama H.;
 RT "Structure and organization of the hepatitis C virus genome isolated
 RT from human carriers";
 RL J. Virol. 65:1105-1113(1991).
 RN [2]
 RP SEQUENCE OF 1487-1500.
 RX MEDLINE=96235224; PubMed=8647104;
 RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetky L.;
 RT "Non-structural protein 3 of hepatitis C virus inhibits

RT phosphorylation mediated by CAMP-dependent protein kinase." ;
 RL Eur. J. Biochem. 237:611-618(1996).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RX MEDLINE=97015088; PubMed=8861916;
 RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
 RA Moonaw E.W., Adachi T., Hostomsky Z.,
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 RT trypsin-like fold and a structural zinc binding site." ;
 RL Cell 87:331-342(1996).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RX MEDLINE=98227846; PubMed=9568891;
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
 RT virus: a 2.2-A resolution structure in a hexagonal crystal form." ;
 RL Protein Sci. 7:837-847(1998).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polypeptide, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA)(N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M58335; AAA72945.1; -.
 DR PIR; A38465; GNMVTC.
 DR PDB; 1AJQ; 25-MAR-98.
 DR PDB; 1JXP; 14-JAN-98.
 DR PDB; 1NS3; 08-APR-98.
 DR PDB; 1C2P; 15-NOV-00.
 DR PDB; 1CSJ; 08-NOV-99.
 DR PDB; 1GX5; 09-APR-02.
 DR PDB; 1GX6; 10-APR-02.
 DR PDB; 1QUV; 26-JUN-00.
 DR PDB; 8OHM; 20-APR-99.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRp.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.

DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 DR INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 DR CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH_BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT STRAND 1031 1035 STRAND
 FT HELIX 1039 1047 STRAND
 FT STRAND 1050 1050 STRAND
 FT STRAND 1059 1063 STRAND
 FT STRAND 1068 1074 STRAND
 FT TURN 1075 1076 STRAND
 FT STRAND 1077 1081 STRAND
 FT HELIX 1082 1085 STRAND
 FT TURN 1086 1087 STRAND
 FT STRAND 1090 1092 STRAND
 FT TURN 1093 1094 STRAND
 FT STRAND 1095 1097 STRAND
 FT STRAND 1101 1103 STRAND
 FT TURN 1104 1107 STRAND
 FT STRAND 1108 1112 STRAND
 FT STRAND 1120 1120 STRAND
 FT STRAND 1122 1122 STRAND
 FT STRAND 1129 1133 STRAND
 FT TURN 1135 1136 STRAND
 FT STRAND 1139 1144 STRAND
 FT STRAND 1149 1157 STRAND
 FT HELIX 1158 1161 STRAND
 FT TURN 1162 1163 STRAND
 FT TURN 1165 1166 STRAND
 FT STRAND 1168 1171 STRAND
 FT TURN 1172 1174 STRAND
 FT STRAND 1175 1186 STRAND

FT TURN 1187 1188
 FT STRAND 1189 1197
 FT HELIX 1198 1202
 FT TURN 1203 1204
 FT STRAND 1680 1688
 SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCFDFD9C CRC64;

Query Match 30.0%; Score 6; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 83;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CGSGPW 8
 |||||
 DB 597 CGSGPW 602

RESULT 10
 POLG_HCVJA
 ID POLG_HCVJA STANDARD; PRT; 3010 AA.
 AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Japanese) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9108850; PubMed=2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 RA Sugimura T., Shimotohno K.;
 RT "Molecular cloning of the human hepatitis C virus genome from
 RT Japanese patients with non-A, non-B hepatitis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RN [2]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=91192160; PubMed=1849488;
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K.,
 RA Ohkoshi S., Shimotohno K.;
 RT "Molecular structure of the Japanese hepatitis C viral genome.";
 RL FEBS Lett. 280:325-328(1991).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; D90208; BAA14233.1; --
 DR FIR; A39253; GNWVCU.

DR HSP; P26663; LXP.
 DR MEROPS; S29.001; --
 DR MEROPS; U39.001; --
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF02907; HCV_NS4a; 1.
 DR Pfam; PF01006; HCV_NS4b; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 83;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CGSGPW 8
 Db 597 CGSGPW 602

RESULT 11
 ID POLG_HCVJT STANDARD; PRT; 3010 AA.
 AC Q00269;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-JT) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31642;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=92295714; PubMed=1318627; Ootsuyama Y., Cho M.J.,
 RA Tanaka T., Kato N., Nakagawa M., Ishimura Y., Shimotohno K.;
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 RT "Molecular cloning of hepatitis C virus genome from a single Japanese
 RT carrier: sequence variation within the same individual and among
 RT infected individuals."
 RL Virus Res. 23:39-53 (1992).

CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA}(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D11168; BAA01943.1; --
 CC PIR; A45573; A45573.
 CC PDB; 1A1Q; 25-MAR-98.
 CC PDB; 1JXP; 14-JAN-98.
 CC MEROPS; S29.001; --
 CC MEROPS; U39.001; --
 CC InterPro; IPR001410; DEAD.
 CC InterPro; IPR002522; HCV_capsid.
 CC InterPro; IPR002521; HCV_core.
 CC InterPro; IPR002519; HCV_env.
 CC InterPro; IPR002531; HCV_NS1.
 CC InterPro; IPR002518; HCV_NS2.
 CC InterPro; IPR004109; HCV_NS3.
 CC InterPro; IPR000745; HCV_NS4.
 CC InterPro; IPR001490; HCV_NS4a.
 CC InterPro; IPR002868; HCV_NS5a.
 CC InterPro; IPR002166; HCV_RdRP.
 CC InterPro; IPR007095; RNA_pol_DS_ps.

DR InterPro; IPR007094; RNA_pol_PSvir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.

FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NON-STRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NON-STRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 2014 3010 NON-STRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT ACT_SITE 1083 1083 POTENTIAL.
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH_BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C7743D5D642BB CRC64;

Query Match 30.0%; Score 6; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 83;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CGSGPW 8
 Db 597 CGSGPW 602

RESULT 12
 ID POLG_HCVTW STANDARD; PRT; 3010 AA.
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Taiwan) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31645;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230206; PubMed=1314449;
RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
RT "The Taiwanese hepatitis C virus genome: sequence determination and
RL mapping the 5' terminus of viral genomic and antigenomic RNA.";
RL Virology 188:102-113(1992)
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA) (N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M84754; -; NOT_ANNOTATED_CDS.
DR PIR: A02444; GNWVTH.
DR FDB: IN64; 25-FEB-03.
DR FDB: IN53; 08-APR-98.
DR MEROPS: S29.001; -.
DR MEROPS: U39.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;

KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CELLULAR AMINOPEPTIDASE.
FT CHAIN 1 115 CORE PROTEIN (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 233 233 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 (N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 (N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;
Query Match 30.0%; Score 6; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 CGSGPW 8
DB 597 CGSGPW 602
RESULT 13
POLG_HCVJ6
ID POLG_HCVJ6 STANDARD; PRT; 3033 AA.
AC P26660;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J6) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11113;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92044440; PubMed=1658196;

Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
Machida A., Miyakawa Y., Mayumi M.;
"Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
from a human carrier: comparison with reported isolates for conserved
and divergent regions.";
J. Gen. Virol. 72:2697-2704(1991).
-!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
-!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
precursor polyprotein, commonly with Asp or Glu in the P6
position, Cys or Thr in P1 and Ser or Ala in P1'.
-!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
{RNA}(n).
-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA.
-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; D00944; BAA00792.1; -;
PIR; J01303; J01303.
HSSP; P27958; IHEI.
DR MEROPS; S29.001; -;
DR MEROPS; U39.001; -;
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PsVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase C; 1.
DR Pfam; PF00998; Viral RdRp; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDc1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
NONSTRUCTURAL PROTEIN NS2A (POTENTIAL).
NONSTRUCTURAL PROTEIN NS2B (POTENTIAL).
MAJOR ENVELOPE PROTEIN E (POTENTIAL).
NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
NONSTRUCTURAL PROTEIN NS3 (POTENTIAL).
PROTEASE/HELICASE NS4 (POTENTIAL).
NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).

FT CHAIN 2018 3033 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT SITE 1087 1087 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 1111 1111 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 1169 1169 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1234 1241 ATP (POTENTIAL).
FT SITE 1320 1323 DECH_BOX.
FT CARBOHYD 136 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;
Query Match 30.0%; Score 6; DB 1; Length 3033;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 CGSGPW 8
DB 601 CGSGPW 606

RESULT 14
CYB_BRANA STANDARD; PRT; 89 AA.
ID CYB_BRANA
AC P49390;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cytochrome b (Fragment).
GN MTCYB OR COB OR CYTB.
OS Brassica napus (Rape).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurasids II; Brassicales; Brassicaceae; Brassica.
OC
OX NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RX MEDLINE=94073988; PubMed=8252643;
RA Ye F., Bernhardt J., Abel W.O.;
RT "Genes for ribosomal proteins S3, L16, L5 and S14 are clustered in
the mitochondrial genome of Brassica napus L.";
RL Curr. Genet. 24:323-329(1993).
CC -!- FUNCTION: Component of the ubiquinol-cytochrome c reductase
complex (complex III or cytochrome b-c1 complex), which is a
respiratory chain that generates an electrochemical potential
coupled to ATP synthesis (By similarity).
CC -!- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or
b562) is low-potential and absorbs at about 562, and heme 2 (or BH
or b566) is high-potential and absorbs at about 566 (By
similarity).
CC -!- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
cytochrome c1 and the Rieske protein (By similarity).
CC -!- SIMILARITY: Belongs to the cytochrome b family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: X68727; CAA48669.1; --
DR InterPro: IPR005798; Cytb_b6_C.
DR InterPro: IPR005797; Cytb_b6_N.
DR Pfam: PF00033; cytochrome b_N; 1.
DR PROSITE: PS00192; CYTOCHROME B_HEME; 1.
DR PROSITE: PS00193; CYTOCHROME B_QO; PARTIAL.
KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW Heme.
FT METAL 88 88 IRON 1 (HEME B562 AXIAL LIGAND).
FT NON_TER 89 89
SQ SEQUENCE 89 AA; 10215 MW; 9924A3BABF6DAB61 CRC64;

Query Match 25.0%; Score 5; DB 1; Length 89;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 LVDYP 18
|||
DB 23 LVDYP 27

RESULT 15
Y001 ARCFU
ID Y001 ARCFU STANDARD; PRT; 126 AA.
AC Q30234;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF0001.
GN AF0001.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: AE001107; AAB91240.1; --
DR PIR: A69250; A69250.
DR TIGR: AF0001; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 126 AA; 14474 MW; A0916AE5D08F298B CRC64;

Query Match 25.0%; Score 5; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 TPRCL 14
|||
DB 102 TPRCL 106

Search completed: November 21, 2003, 21:00:01
Job time : 5.85 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:45:57 ; Search time 24.4 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-87

Perfect score: 20
Sequence: 1 SRCGSGPWITPRCLVDYDYPYR 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	137	Q914V3	Q914V3 hepatitis c
2	20	100.0	137	Q914V1	Q914V1 hepatitis c
3	20	100.0	137	Q914V6	Q914V6 hepatitis c
4	20	100.0	137	Q914V4	Q914V4 hepatitis c
5	20	100.0	137	Q914V0	Q914V0 hepatitis c
6	20	100.0	137	Q914V5	Q914V5 hepatitis c
7	20	100.0	137	Q914V2	Q914V2 hepatitis c
8	20	100.0	137	Q914U9	Q914U9 hepatitis c
9	20	100.0	420	Q98UN4	Q98UN4 hepatitis c
10	20	100.0	640	Q68966	Q68966 hepatitis c
11	20	100.0	746	Q8JPM2	Q8JPM2 hepatitis c
12	20	100.0	2436	Q81756	Q81756 hepatitis c
13	20	100.0	3011	Q81FE5	Q81FE5 hepatitis c
14	20	100.0	3011	Q83463	Q83463 hepatitis c
15	20	100.0	3011	Q9DIT6	Q9DIT6 hepatitis c
16	15	75.0	128	Q8JYQ3	Q8JYQ3 hepatitis c

17	15	75.0	129	12	Q8JYQ1	Q8JYQ1 hepatitis c
18	15	75.0	129	12	Q8JYQ9	Q8JYQ9 hepatitis c
19	15	75.0	129	12	Q8JYQ0	Q8JYQ0 hepatitis c
20	15	75.0	137	12	Q914Q6	Q914Q6 hepatitis c
21	15	75.0	137	12	Q914W1	Q914W1 hepatitis c
22	15	75.0	137	12	Q914Q5	Q914Q5 hepatitis c
23	15	75.0	137	12	Q914Q8	Q914Q8 hepatitis c
24	15	75.0	137	12	Q914W4	Q914W4 hepatitis c
25	15	75.0	137	12	Q914W9	Q914W9 hepatitis c
26	15	75.0	137	12	Q914W2	Q914W2 hepatitis c
27	15	75.0	137	12	Q914W8	Q914W8 hepatitis c
28	15	75.0	137	12	Q914Q4	Q914Q4 hepatitis c
29	15	75.0	137	12	Q914W5	Q914W5 hepatitis c
30	15	75.0	137	12	Q914X0	Q914X0 hepatitis c
31	15	75.0	137	12	Q914V9	Q914V9 hepatitis c
32	15	75.0	137	12	Q914V8	Q914V8 hepatitis c
33	15	75.0	137	12	Q914W6	Q914W6 hepatitis c
34	15	75.0	137	12	Q914Q2	Q914Q2 hepatitis c
35	15	75.0	137	12	Q914Q1	Q914Q1 hepatitis c
36	15	75.0	137	12	Q914W7	Q914W7 hepatitis c
37	15	75.0	137	12	Q914Q7	Q914Q7 hepatitis c
38	15	75.0	137	12	Q914V7	Q914V7 hepatitis c
39	15	75.0	137	12	Q914W3	Q914W3 hepatitis c
40	15	75.0	403	12	Q9PX22	Q9PX22 hepatitis c
41	14	70.0	137	12	Q914W0	Q914W0 hepatitis c
42	13	65.0	128	12	Q8JYQ2	Q8JYQ2 hepatitis c
43	13	65.0	778	12	Q04185	Q04185 hepatitis c
44	13	65.0	778	12	Q04184	Q04184 hepatitis c
45	13	65.0	3011	12	Q36579	Q36579 hepatitis c

ALIGNMENTS

RESULT 1
Q914V3 ID Q914V3 PRELIMINARY; PRT; 137 AA.
AC Q914V3;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Genome polyprotein (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RC STRAIN=sh;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before and after liver transplantation".
RL Submitted (OCT-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF431834; AAL30684.1; -;
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 15024 MW; E374842A36F3E2E7 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 3.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYDYPYR 20
Db 117 SRCGSGPWITPRCLVDYDYPYR 136

RESULT 2

```

Q914V1
ID Q914V1 PRELIMINARY; PRT; 137 AA.
AC Q914V1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=sh;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431836; AAL30686.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR PolyProtein; Transmembrane.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 14994 MW; FEA94437FB33E2E7 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 3.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRCGSGPWITPRCLVDYPYR 20
Db 117 SRCGSGPWITPRCLVDYPYR 136

RESULT 3
Q914V6
ID Q914V6 PRELIMINARY; PRT; 137 AA.
AC Q914V6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=sh;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431831; AAL30681.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR PolyProtein; Transmembrane.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 14966 MW; 5EB0443D75188F51 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 3.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRCGSGPWITPRCLVDYPYR 20
Db 117 SRCGSGPWITPRCLVDYPYR 136

Q914V4
ID Q914V4 PRELIMINARY; PRT; 137 AA.
AC Q914V4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=sh;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431833; AAL30683.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR PolyProtein; Transmembrane.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 14994 MW; FEA94437FB33E2E7 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 3.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRCGSGPWITPRCLVDYPYR 20
Db 117 SRCGSGPWITPRCLVDYPYR 136

RESULT 5
Q914V0
ID Q914V0 PRELIMINARY; PRT; 137 AA.
AC Q914V0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=sh;
RA Lyra A.C., Fan X., Di Bisceglie A.M.;
RT "Genetic diversity of two putative cd81 binding regions of HCV before
and after liver transplantation.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF431837; AAL30687.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR PolyProtein; Transmembrane.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 137
SQ SEQUENCE 137 AA; 14994 MW; FEA94437FB33E2E7 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 3.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRCGSGPWITPRCLVDYPYR 20
Db 117 SRCGSGPWITPRCLVDYPYR 136

```


QY 1 SRCGSGPWITPRCLVDYPYR 20
 Db 117 SRCGSGPWITPRCLVDYPYR 136

RESULT 6

Q914V5 ID Q914V5 PRELIMINARY; PRT; 137 AA.
 AC Q914V5;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=sh;
 RA Lyra A.C., Fan X., Di Bisceglie A.M.;
 RT "Genetic diversity of two putative cd81 binding regions of HCV before
 and after liver transplantation."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL, AF431832; AAL30682.1; --
 DR InterPro: IPR002531; HCV NS1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR ProDom: PD186062; HCV NS1; 1.
 DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 137
 SQ SEQUENCE 137 AA; 14994 MW; FE94437FB33E2E7 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;

Best Local Similarity 100.0%; Pred. No. 3.9e-15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
 Db 117 SRCGSGPWITPRCLVDYPYR 136

RESULT 7

Q914V2 ID Q914V2 PRELIMINARY; PRT; 137 AA.
 AC Q914V2;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=sh;
 RA Lyra A.C., Fan X., Di Bisceglie A.M.;
 RT "Genetic diversity of two putative cd81 binding regions of HCV before
 and after liver transplantation."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL, AF431835; AAL30685.1; --
 DR InterPro: IPR002531; HCV NS1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR ProDom: PD186062; HCV NS1; 1.
 DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 137
 SQ SEQUENCE 137 AA; 14994 MW; FE94437FB33E2E7 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;

Best Local Similarity 100.0%; Pred. No. 3.9e-15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
 Db 117 SRCGSGPWITPRCLVDYPYR 136

RESULT 8

Q914U9 ID Q914U9 PRELIMINARY; PRT; 137 AA.
 AC Q914U9;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=sh;
 RA Lyra A.C., Fan X., Di Bisceglie A.M.;
 RT "Genetic diversity of two putative cd81 binding regions of HCV before
 and after liver transplantation."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL, AF431838; AAL30688.1; --
 DR InterPro: IPR002531; HCV NS1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR ProDom: PD186062; HCV NS1; 1.
 DR Coa protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 137
 SQ SEQUENCE 137 AA; 15047 MW; E3023437FB33F6A2 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 137;

Best Local Similarity 100.0%; Pred. No. 3.9e-15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRCGSGPWITPRCLVDYPYR 20
 Db 117 SRCGSGPWITPRCLVDYPYR 136

RESULT 9

Q98UN4 ID Q98UN4 PRELIMINARY; PRT; 420 AA.
 AC Q98UN4;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Qla;
 RX MEDLINE=20499063; PubMed=11044085;
 RA Hadlock K.G., Lanford R.E., Perkins S., Rowe J., Yang Q., Levy S.,
 RA Pileri P., Abrignani S., Fount S.K.;
 RT "Human monoclonal antibodies that inhibit binding of hepatitis C virus
 E2 protein to CD81 and recognize conserved conformational epitopes."
 RL J. Virol. 74:10407-10416 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Qla;
 RA Keck Z.Y., Hadlock K.G., Yang Q.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL, AF348704; AAK32685.1; --

```

DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT CHAIN 1 1 E1.
FT CHAIN 38 >400 E2.
FT CHAIN 401 >420 P7.
FT CHAIN 420 420
FT NON TER 420 420
SQ SEQUENCE 420 AA; 45882 MW; F28D0C3EE0CD081E CRC64;

Query Match 100.0%; Score 20; DB 12; Length 420;
Best Local Similarity 100.0%; Pred. No. 9.8e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRCGSGPWITPRCLVDYPYR 20
Db 249 SRCGSGPWITPRCLVDYPYR 268

RESULT 10
Q68966 PRELIMINARY; PRT; 640 AA.
AC Q68966;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Oxf protein (Fragment).
GN ORF.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Kumar U.; Cheng D.; Thomas H.; Monjardino J.;
RX MEDLINE=92300349; PubMed=1318944;
RT "Cloning and sequencing of the structural region and expression of
RT putative core gene of hepatitis C virus from a British case of chronic
RT sporadic hepatitis.";
RL J. Gen. Virol. 73:1521-1525(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA Kumar U.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X84079; CAA58888.1; -.
DR InterPro; IPR002531; HCV NS1.
DR ProDom; PD186062; HCV NS1; 1.
FT CHAIN 1 >191 CORE.
FT CHAIN 192 >383 ENVELOPE 1.
FT CHAIN 384 >640 ENVELOPE 2.
FT NON TER 640 640
SQ SEQUENCE 640 AA; 69857 MW; 760FEC3DE7741959 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 640;
Best Local Similarity 100.0%; Pred. No. 1.4e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRCGSGPWITPRCLVDYPYR 20
Db 595 SRCGSGPWITPRCLVDYPYR 614

RESULT 11
Q8JPM2 PRELIMINARY; PRT; 746 AA.
AC Q8JPM2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)

```

```

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21904749; PubMed=11907246;
RA Blanchard E.; Brand D.; Trassard S.; Goudeau A.; Roingeard P.;
RT "Hepatitis C virus-like particle morphogenesis.";
RL J. Virol. 76:4073-4079(2002).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AF529293; AAM94419.1; -.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR Polyprotein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Coat protein; Transmembrane.
FT NON TER 746 746
SQ SEQUENCE 746 AA; 82116 MW; 2DA94CC6845BA92C CRC64;

Query Match 100.0%; Score 20; DB 12; Length 746;
Best Local Similarity 100.0%; Pred. No. 1.6e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SRCGSGPWITPRCLVDYPYR 20
Db 595 SRCGSGPWITPRCLVDYPYR 614

RESULT 12
Q81756 PRELIMINARY; PRT; 2436 AA.
AC Q81756;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Choo Q.-L.; Richman K.; Han J.;
RT "The nucleotide sequence of the Hepatitis C viral genome.";
RL Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.
DR EMBL; M32084; AAA45677.1; -.
DR HSPSP; P27958; 1A1V
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV NS5b.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.

```

Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase C; 1.
DR Pfam; PF00998; Viral RdRp; 1.
DR PRODOM; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR POSITE; PS00190; CYTOCHROME C; 1.
DR POSITE; PS0507; RDRP POSITIVE; 1.
DR POSITE; PS0521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327124 MW; 2489CE74AC864E58 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred.No. 4.8e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 1 SRCGGPWITPRCLVDYPVR 20
|||||
DB 595 SRCGGPWITPRCLVDYPVR 614
|||

RESULT 14

Q03463 ID Q03463 PRELIMINARY; PRT; 3011 AA.

AT Q03463; AT Q03463; TREMBLrel_01, Created)

DT 01-NOV-1996 (TREMBLrel_01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel_23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX Hepacivirus.
NCBI_TaxId=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=G2044440; PubMed=1658196;
RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A., Miyakawa Y., Mayumi M.;
RA "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions.";
RL Jpn. J. Exp. Med. 60:167-177(1990).
RZ [2]

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=G2044440; PubMed=1658196;
RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A., Miyakawa Y., Mayumi M.;
RA "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions.";
RL J. Gen. Virol. 72:2697-2704(1991).
RZ [3]

RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=G3117120; PubMed=1335573;
RA Okamoto H., Kanai N., Mishiro S.;
RA "Full-length nucleotide sequence of a Japanese hepatitis C virus isolate (HC-J1) with high homology to USA isolates.";
RL Nucleic Acids Res. 20:6410-6410(1992).
RZ [4]

RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=G4174722; PubMed=7510436;
RA Mink M., Benichou S., Madale P., Tiollais P., Prince A.;
RA Submitted (DEC-1992) to the EMBL/GenBank/DDBJ databases.
RZ [5]

RA Inchauspe G.;
 RT "Characterization and mapping of a B-cell immunogenic domain in
 RT Hepatitis C virus E2 glycoprotein using a yeast peptide library.";
 RL Virology 200;246-255(1994).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL; D10749; BAA01582.1; -.
 DR HSP; P27958; 1AIV.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_core; 1.
 DR Pfam; PF01542; HCV env; 1.
 DR Pfam; PF01539; HCV NS1; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; helicase C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS50507; RDRP_POSITIVE; 1.
 DR PROSITE; PS50521; RDRP_VIRAL; 1.
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydroxylase; Nonstructural protein; Polyprotein;
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3011 AA; 327112 MW; 97E9052C0250463B CRC64;
 Query Match 100.0%; Score 20; DB 12; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 4.8e-14;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SRCSGSPWITPRCLVDYPYR 20
 DB 595 SRCSGSPWITPRCLVDYPYR 614
 RESULT 15
 Q9DIT6 PRELIMINARY; PRT; 3011 AA.
 AC Q9DIT6
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Genomic RNA for polyprotein gene (Genome polyprotein).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21014672; PubMed=11115058;
 RA Kumar U., Tuthill T., Thomas H.C., Monjardino J.;
 RT "Sequence, expression and reconstitution of an HCV genome from a
 RT British isolate derived from a single blood donation.";
 RL J. Viral Hepat. 7:459-465(2000).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL; AJ278830; CAC03609.1; -.
 DR HSP; P27958; 1AIV.
 DR InterPro; IPR000345; CytC_heme_bind.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_core; 1.
 DR Pfam; PF01542; HCV env; 1.
 DR Pfam; PF01539; HCV NS1; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; helicase C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS50507; CYTOCHROME C; 1.
 DR PROSITE; PS50507; RDRP_POSITIVE; 1.
 DR PROSITE; PS50521; RDRP_VIRAL; 1.
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydroxylase; Nonstructural protein; Polyprotein;
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3011 AA; 327405 MW; 7B6264A74A5452D3 CRC64;
 Query Match 100.0%; Score 20; DB 12; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 4.8e-14;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SRCSGSPWITPRCLVDYPYR 20
 DB 595 SRCSGSPWITPRCLVDYPYR 614

Search completed: November 21, 2003, 21:08:21
 Job time : 25.45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:22 ; Search time 32.1 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-88

Perfect score: 20

Sequence: 1 CLVDYPRLWHYPCINITYI 20

Scoring table:

Gapol 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	16	AA884493
2	20	100.0	20	17	AA91011
3	20	100.0	20	23	AAO18717
4	20	100.0	21	14	AA833879
5	20	100.0	21	14	AA832419
6	20	100.0	30	23	AAU84638
7	20	100.0	121	14	AA832418
8	20	100.0	179	21	AA818529
9	20	100.0	192	19	AAW67009
					Hepatitis C virus
					HCV E2 peptide E2-
					Hepatitis C virus
					Polypeptide p607 c
					Sequence of immuno
					HCV HepC1a segment
					Sequence of immuno
					Protein encoded by
					HCV nucleocapsid c

10	20	100.0	254	22	AA868043	Amino acid sequenc
11	20	100.0	363	22	AA868042	Amino acid sequenc
12	20	100.0	402	14	AA834438	Sequence of glycop
13	20	100.0	454	10	AA890183	Sequence of hepati
14	20	100.0	454	10	AA892049	Sequence encoded b
15	20	100.0	454	21	AA818526	Protein encoded by
16	20	100.0	480	14	AA833992	HCV-1 E2/NS1 prote
17	20	100.0	531	22	AAE02622	Chimeric HCV E2661
18	20	100.0	621	14	AA833185	Sequence of subtra
19	20	100.0	622	14	AA833591	HCV CKS-NS1S1-NS1S
20	20	100.0	622	22	AA869010	HCV recombinant an
21	20	100.0	637	24	ABP57410	Hepatitis C virus
22	20	100.0	663	17	AA892935	HCV E2 + NS2 poly
23	20	100.0	663	20	AAW67615	Hepatitis C virus
24	20	100.0	733	14	AA838278	NANB hepatitis vir
25	20	100.0	738	14	AA833592	HCV CKS-full lengt
26	20	100.0	738	22	AA869011	HCV recombinant an
27	20	100.0	1997	23	AAU84802	HCV HepC cassette
28	20	100.0	2435	13	AA825135	HCV polypeptide 1
29	20	100.0	2436	10	AA892050	Sequence encoded i
30	20	100.0	2436	10	AA890288	Peptide encoded by
31	20	100.0	2436	13	AA828582	HCV amino acid seq
32	20	100.0	2772	11	AA808123	Hepatitis C virus
33	20	100.0	2772	21	AA818540	Protein encoded by
34	20	100.0	2816	14	AA834009	HCV-1 polypeptid
35	20	100.0	2894	13	AA824440	Composite HCV HC-J
36	20	100.0	2894	16	AA870230	Hepatitis C virus
37	20	100.0	2955	11	AA808124	Amino acid sequenc
38	20	100.0	2955	20	AA814975	Polypeptide encode
39	20	100.0	2955	21	AA818541	Compiled HCV seque
40	20	100.0	3011	13	AA821519	Hepatitis C virus
41	20	100.0	3011	14	AA831621	Hepatitis C virus
42	20	100.0	3011	15	AA865995	Hepatitis C virus
43	20	100.0	3011	17	AA890931	Hepatitis C virus
44	20	100.0	3011	18	AAW34480	HCV polypeptid
45	20	100.0	3011	19	AAW40038	HCV polypeptid

ALIGNMENTS

RESULT 1
AA884493
ID AA884493 standard; peptide; 20 AA.
XX AC
AA884493;
XX AC
DT 06-JAN-1997 (first entry)
XX
DE Hepatitis C virus T-cell epitope peptide NS1-27 (residues 607-626).
XX
KW Hepatitis C virus; HCV; immunogen; non-structural region;
KW immunodominant; T cell epitope; vaccine.
XX
OS Hepatitis C virus.
XX
PN WO9512677-A2.
XX
PD 11-MAY-1995.
XX
PF 28-OCT-1994; 94WO-EP03555.
XX
PR 04-NOV-1993; 93EP-0402718.
XX
XX (INNO-) INNOGENETICS NV.
XX
PI Deleys R, Leroux-Roels G, Maertens G;
XX
DR WPI; 1995-193822/25.
XX
PT Hepatitis C Virus immunogenic polypeptide contg. a T-cell
PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in
PT production of vaccines, therapeutic agents, etc.

XX PS Claim 26; Page 71; 105pp; English.

XX CC The present sequence is a specifically claimed example of a

CC CC T-cell epitope-containing peptide derived from hepatitis C virus.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 16; Length 20;

Best Local Similarity 100.0%; Pred. No. 4.8e-16;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPTINYTI 20

Db 1 CLVDYPYRLWHYPTINYTI 20

RESULT 2

AA018717 standard; peptide; 20 AA.

AC AAO18717;

DT 25-SEP-1996 (first entry)

DE HCV E2 peptide E2-27 for competition studies.

XX HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;

KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.

XX Synthetic.

XX WO9604385-A2.

PD 15-FEB-1996.

XX 31-JUL-1995; 95WO-EP03031.

XX 29-JUL-1994; 94EP-0870132.

XX (INNO-) INNOGENETICS NV.

PI Bosman F, Buyse M, De Martynoff G, Maertens G;

PI WPI; 1996-129401/13.

XX Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope

PT proteins - in presence of disulphide bond cleavage agent, to

PT produce proteins suitable for direct use in vaccines or diagnostic

PT assays of HCV

XX Claim 29; Page 67; 146pp; English.

XX AAR90965-R90980 and AAR90995-R91015 represent synthetic hepatitis C

CC virus (HCV) E1 and E2 peptides used in competition studies. This

CC sequence represents a synthetic E2 peptide, and corresponds to residues

CC 607-626 of the E2 protein sequence. These sequences are useful for in

CC vitro monitoring of HCV disease, or prognosis of the response to

CC interferon treatment of patients suffering from HCV infection. These

CC sequences compete with the proteins produced by AAT12704-T12709 and

CC AAT12961-T12974, which are included in vectors for the production of

CC recombinant E1, E2, and E1/E2. The recombinant proteins can then be

CC isolated and purified by carrying out a disulphide bond cleavage, or a

CC reduction step with a disulphide bond cleavage agent, after lysis of

CC recombinant host cells. The constructs containing the purified HCV

CC envelope proteins can be used for vaccinating humans against HCV, for in

CC vitro detection of HCV antibodies in a sample, and in a serotyping assay

CC for detecting one or more serological types of HCV present in a

CC biological sample. The constructs can also be immobilised on a solid

CC substrate and incorporated into a reversed phase hybridisation assay for

CC determining the presence or the genotype of HCV. The new purification

CC method preserves the conformation of the recombinantly expressed E1, E2

CC and E1/E2, and eliminates contaminating proteins. Antigens isolated

CC using this method are more reactive with human sera than those isolated

CC by known techniques.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 17; Length 20;

Best Local Similarity 100.0%; Pred. No. 4.8e-16;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPTINYTI 20

Db 1 CLVDYPYRLWHYPTINYTI 20

RESULT 3

AA018717 standard; Peptide; 20 AA.

XX AAO18717;

DT 24-OCT-2002 (first entry)

DE Hepatitis C virus E2 protein derived peptide E2-27.

XX Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide;

KW immunostimulant; vaccine.

XX Hepatitis C virus.

XX WO200255548-A2.

XX 18-JUL-2002.

XX 11-JAN-2002; 2002WO-EP00219.

XX 11-JAN-2001; 2001US-260699P.

PR 30-AUG-2001; 2001US-315768P.

XX (INNO-) INNOGENETICS NV.

XX Maertens G, Bosman F, Buyse M;

XX WPI; 2002-599657/64.

XX New therapeutic vaccine compositions comprising at least one purified

PT recombinant hepatitis C virus (HCV) single or specific oligomeric

PT recombinant envelope protein E1 or E2, useful for immunizing humans

PT from HCV infection

XX Claim 4; Page 230; 243pp; English.

XX The present invention relates to new therapeutic vaccine compositions for

CC inducing hepatitis C virus (HCV)-specific antibodies, comprising a

CC composition containing at least one purified recombinant HCV single or

CC specific oligomeric recombinant envelope proteins selected from an E1 and

CC an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are

CC useful for inducing HCV-specific antibodies or for immunising humans

CC against HCV. The recombinant HCV E1 and/or E2 proteins are useful as

CC vaccines or therapeutics, in HCV screening and confirmatory antibody

CC tests, for raising antibodies, in the preparation of medicament, and for

CC in vitro monitoring of HCV disease or prognosis of the response to

CC treatment of patients suffering from HCV infection. The present sequence

CC is a peptide derived from the proteins of the invention.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 23; Length 20;

Best Local Similarity 100.0%; Pred. No. 4.8e-16;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPTINYTI 20

Db 1 CLVDYPYRLWHYPTINYTI 20

```

XX 04-MAR-1993.
XX
XX PD
XX XX
XX PF 21-AUG-1992; 92WO-US07189.
XX XX
XX PR 21-AUG-1991; 91US-0748292.
XX XX
XX PA (ABBO ) ABBOTT LAB.
XX XX
XX PI Dailey SH, Desai SM, Devare SG, Johnson JE, Mehta SU;
XX DR WPI; 1993-094030/11.
XX XX
XX PT Monoclonal antibodies specific for hepatitis C virus EZ-NS1
XX PT antigen - useful for diagnosis and evaluation of HCV infections
XX PT and in differentiation studies
XX XX
XX PS Example; Page 30; 48pp; English.
XX XX
XX CC Immunogenic domains of E2/NS1 region of HCV genome encompassing AAs
XX CC 600-720 were mapped with PEPSCAN analysis. FAB dimers of IgG
XX CC purified from sera of individuals seropositive for antibodies to HCV
XX CC proteins were used as the primary antibody for the serological
XX CC analysis. Based on the reactivity these sera in EIA, four AA
XX CC sequences (AAR32419,R32420,AAR33182,R33183) were identified as
XX CC immunogenic domains. Each of these four sequences and an additional
XX CC sequence, which was the combination of the two most immunogenic
XX CC sequences (AAR33184) were synthesised. Based on the EIA reactivity of
XX CC a panel of HCV positive sera, peptide AAR33184 was chosen as the
XX CC immunogen for the generation of monoclonal antibodies to HCV NS1.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX XX
XX SQ Sequence 21 AA;

Query Match 100.0%; Score 20; DB 14; Length 21;
Best Local Similarity 100.0%; Pred. No. 5e-16; 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0;

Qy 1 CLVDYPYRLWHYPCTINYTI 20
   |||||
Db 1 CLVDYPYRLWHYPCTINYTI 20

RESULT 6
AAU84638
ID AAU84638 standard; Peptide; 30 AA.
XX
XX AC AAU84638;
XX
XX DT 08-MAY-2002 (first entry)
XX
XX DE HCV HepC1a segment 41.
XX
XX KW Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
XX KW viral infection; human immunodeficiency virus; melanoma;
XX KW bacterial infection; Salmonella; Legionella; parasitic infection;
XX KW Trypanosoma; Toxoplasma; Giardia.
XX
XX OS Hepatitis C virus.
XX
XX PN WO200190197-A1.
XX
XX PD 29-NOV-2001.
XX
XX PF 25-MAY-2001; 2001WO-AU00622.
XX
XX PR 25-MAY-2000; 2000AU-0007761.
XX
XX PA (AUSU ) UNIV AUSTRALIAN NAT.
XX
XX PI Thomson SA, Ramshaw IA;
XX DR WPI; 2002-147575/19.

```

DR N-PSDB; ABK36476.
 XX
 PT New synthetic polypeptides having several different segments of at
 PT least one parent polypeptide linked together differently compared to
 PT the linkage in the parent polypeptide, for inducing immune response
 PT against a pathogen or cancer
 XX
 XX Example 2; Fig 26; 364pp; English.
 PS
 CC The invention relates to a new synthetic polypeptide (I) comprising
 CC several different segments of at least one parent polypeptide linked
 CC together in a different relationship relative to their linkage in the
 CC parent polypeptide to impede, abrogate or otherwise alter at least one
 CC function associated with the parent polypeptide and for inducing an
 CC immune response against a pathogen or cancer. Also included are a
 CC synthetic polynucleotide encoding and a computer system for
 CC designing the synthetic polypeptides. The synthetic polypeptides and
 CC polynucleotides are referred to as a vaccine. The synthetic polypeptide is
 CC useful for modulating immune responses preferably directed against a
 CC pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix,
 CC colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone
 CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.
 CC Compositions comprising the polypeptide may be used in the treatment or
 CC prophylaxis against viral (such as infections caused by HIV (human
 CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
 CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
 CC (e.g., infections caused by Neisseria, Meningococcus, Haemophilus,
 CC Salmonella, Streptococcus, Legionella and Mycobacterium) or parasitic
 CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
 CC Trypanosoma, Toxoplasma and Giardia) infections. The present
 CC sequence is a peptide derived from a parent protein used to
 CC construct a vaccine of the invention.
 XX
 SQ Sequence 30 AA;
 Query Match 100.0%; Score 20; DB 23; Length 30;
 Best Local Similarity 100.0%; Pred. No. 6.8e-16;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CLVDYPYRLWHYPTINVTI 20
 |||||
 Db 9 CLVDYPYRLWHYPTINVTI 28
 |||||
 RESULT 7
 AAR32418
 ID AAR32418 standard; peptide; 121 AA.
 XX
 AC AAR32418;
 XX
 XX 25-MAR-2003 (updated)
 DT 03-JUL-1993 (first entry)
 XX
 XX Sequence of immunogenic domain of E2/NS1 region of hepatitis C virus
 DE (HCV) genome encompassing AAs 600-720.
 XX
 XX Immunogenic peptide; hepatitis C virus; immunogenic domain;
 KW monoclonal antibody; diagnosis; detection; therapy.
 XX
 OS Synthetic.
 XX
 XX WO9304205-A1.
 PN
 XX
 XX 04-MAR-1993.
 PD
 XX
 XX 21-AUG-1992; 92WO-US07189.
 PF
 XX
 XX 21-AUG-1991; 91US-0748292.
 PR
 XX (ABBO) ABBOTT LAB.
 PA
 XX Dailey SH, Desai SM, Devare SG, Johnson JE, Mehta SU;
 PI
 XX

DR WPI; 1993-094030/11.
 XX
 PT Monoclonal antibodies specific for hepatitis C virus E2-NS1
 PT antigen - useful for diagnosis and evaluation of HCV infections
 PT and in differentiation studies
 XX
 XX Example; Page 29; 48pp; English.
 PS
 CC Immunogenic domains of E2/NS1 region of HCV genome encompassing AAs
 CC 600-720 were mapped with PRPSCAN analysis. FAB dimers of IgG
 CC purified from sera of individuals seropositive for antibodies to HCV
 CC proteins were used as the primary antibody for the serological
 CC analysis. Based on the reactivity these sera in EIA, four AA
 CC sequences (AAR32418, R32420, AAR33182, R33183) were identified as
 CC immunogenic domains. Each of these four sequences and an additional
 CC sequence, which was the combination of the two most immunogenic
 CC sequences (AAR33184) were synthesised. Based on the EIA reactivity of
 CC a panel of HCV positive sera, peptide AAR33184 was chosen as the
 CC immunogen for the generation of monoclonal antibodies to HCV NS1.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 121 AA;
 Query Match 100.0%; Score 20; DB 14; Length 121;
 Best Local Similarity 100.0%; Pred. No. 2.3e-15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CLVDYPYRLWHYPTINVTI 20
 |||||
 Db 8 CLVDYPYRLWHYPTINVTI 27
 |||||
 RESULT 8
 AAB18529
 ID AAB18529 standard; Protein; 179 AA.
 XX
 AC AAB18529;
 XX
 XX 15-JAN-2001 (first entry)
 DT
 XX Protein encoded by a novel hepatitis C virus cDNA clone 131.
 XX Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
 KW viral infectivity; viral replication.
 XX
 OS Hepatitis C virus.
 XX
 XX EP1034785-A2.
 PN
 XX
 XX 13-SEP-2000.
 PD
 XX
 XX 16-MAR-1990; 2000EP-0109602.
 PF
 XX
 XX 17-MAR-1989; 89US-0325338.
 PR
 XX 20-APR-1989; 89US-0341334.
 PR
 XX 18-MAY-1989; 89US-0355002.
 PR
 XX 16-MAR-1990; 90EP-0302866.
 PR
 XX (CHIR) CHIRON CORP.
 PA
 XX Houghton M, Choo Q, Kuo G;
 PI
 XX WPI; 2000-566891/53.
 DR
 DR N-PSDB; AAR75285.
 DR
 XX Novel composition comprising a hepatitis C virus antisense
 PT polynucleotide which is complementary to or corresponds to a sense
 PT strand of the virus genome, and selectively hybridises to it -
 XX
 XX Example; Fig 5; 75pp; English.
 PS
 XX The specification describes a pharmaceutical composition which
 CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The

CC HCV is characterized by a positive stranded RNA genome which has
 CC 40% homology at the polypeptide level to a HCV polyprotein. The
 CC antisense polynucleotide binds to cellular polynucleotides which
 CC enhance and/or are required for viral infectivity, replicative
 CC ability or chronicity. The antisense polynucleotides may also be
 CC designed to bind with high specificity, to be of increased stability,
 CC to be stable and to have low toxicity. The composition also comprises
 CC an agent which causes viral RNA to be inactive. The composition
 CC is used for preventing HCV replication in a system. The present
 CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
 CC course of the invention.

XX SQ Sequence 179 AA;

Query Match 100.0%; Score 20; DB 21; Length 179;
 Best Local Similarity 100.0%; Pred. No. 3.2e-15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
 DB 97 CLVDYPYRLWHYPCTINYTI 116

RESULT 9

AAW67009
 ID AAW67009 standard; protein; 192 AA.

XX AC AAW67009;

DT 02-MAR-1999 (first entry)

XX DE HCV nucleocapsid core protein.

XX KW Hepatitis C virus; HCV; nucleocapsid; core protein; analogue; antibody;
 XX non-structural protein; thioamide bond; peptide bond.

XX OS Hepatitis C virus.

XX FH Key Location/Qualifiers
 FT Misc-difference 79 /note= "given in specification as Try"

XX PN JP10226698-A.

XX PD 25-AUG-1998.

XX PF 19-FEB-1997; 97JP-0034702.

XX PR 19-FEB-1997; 97JP-0034702.

XX PA (KYOW) KYOWA MEDEX KK.

XX DR WPI; 1998-515103/44.

XX PT Determination of antibody in sample - uses peptide analog absorbed
 XX or chemically bound on carrier as antigen

XX PS Disclosure; Page 4; 13pp; Japanese.

XX CC This sequence represents the Hepatitis C virus (HCV) nucleocapsid core
 CC protein. The invention relates to peptide analogues derived from HCV
 CC proteins, e.g. AAW67417-W67426, which can be used for the determination
 CC of anti-HCV antibodies in a sample. Preferably the peptide analogues
 CC contain one or more thioamide peptide bonds where at least one oxygen
 CC atom of the peptide bond is replaced by sulphur atom. The peptide
 CC analogues can be adsorbed or chemically bound to a carrier.

XX SQ Sequence 192 AA;

Query Match 100.0%; Score 20; DB 19; Length 192;
 Best Local Similarity 100.0%; Pred. No. 3.3e-15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
 DB 157 CLVDYPYRLWHYPCTINYTI 176

RESULT 10

AAW68043
 ID AAW68043 standard; protein; 254 AA.

XX AC AAW68043;

XX DT 29-JUN-2001 (first entry)

XX DE Amino acid sequence of water soluble variant of envelope E2 protein.
 XX KW Eo protein; HCV; envelope protein; E2 protein; HCV infection;
 XX HCV attachment.

XX OS Synthetic.

XX OS Hepatitis C virus.

XX PN WO200122984-A1.

XX PD 05-APR-2001.

XX PF 26-SEP-2000; 2000WO-US26395.

XX PR 29-SEP-1999; 99US-0407430.

XX PA (UYCO) UNIV COLUMBIA NEW YORK.

XX PI Worman HJ, Mamiya N;

XX DR WPI; 2001-273486/28.

XX PT Treating or preventing hepatitis C virus infection in a subject,
 XX involves administering hepatitis C virus envelope protein E2 binding
 XX agents

XX PS Claim 5; Fig 8; 46pp; English.

XX CC The present sequence represents a water soluble variant of a Hepatitis C
 CC virus (HCV) envelope E2 protein. Agents that bind to the HCV E2 protein
 CC (such as the human Eo protein), and so inhibit the attachment of HCV onto
 CC cells (especially liver cells), are used to treat HCV infections in
 CC mammals, in particular humans. The specification also describes a method
 CC for identifying a compound which can be used for treating or preventing
 CC HCV in a subject and which can inhibit the attachment of HCV onto cells
 CC by inhibiting the binding of HCV envelope E2 protein to a cellular
 CC protein associated with HCV attachment and entry into cells. The method
 CC comprises incubating the compound, HCV envelope E2 protein or its variant
 CC and a cellular protein capable of specifically binding to the HCV E2
 CC protein under suitable reaction conditions; determining the interactions
 CC between HCV envelope E2 protein and cellular protein in the presence and
 CC absence of the compound; and comparing the interaction to identify a
 CC compound which can inhibit the attachment of HCV onto cells.

XX SQ Sequence 254 AA;

Query Match 100.0%; Score 20; DB 22; Length 254;
 Best Local Similarity 100.0%; Pred. No. 4.3e-15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
 DB 202 CLVDYPYRLWHYPCTINYTI 221

RESULT 11

AAW68042
 ID AAW68042 standard; protein; 363 AA.

XX AC AAW68042;

XX 29-JUN-2001 (first entry)
 DT
 XX Amino acid sequence of a Hepatitis C virus envelope E2 protein.
 DE
 XX E2 protein; HCV; envelope protein; E2 protein; HCV infection;
 KW HCV attachment.
 XX
 OS Hepatitis C virus.
 XX
 XX WO200122984-A1.
 PN
 XX 05-APR-2001.
 PD
 XX 26-SEP-2000; 2000WO-US26395.
 PF
 XX 29-SEP-1999; 99US-0407430.
 PR
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 PA
 XX Worman HJ, Mamiya N;
 PI
 XX WPI; 2001-273486/28.
 DR
 XX Treating or preventing hepatitis C virus infection in a subject,
 XX PT involves administering hepatitis C virus envelope protein E2 binding
 XX PT agents -
 XX
 PS Claim 3; Fig 7; 46pp; English.
 XX
 CC The present sequence represents a Hepatitis C virus (HCV) envelope
 CC E2 protein. Agents that bind to the HCV E2 protein (such as the human
 CC E2 protein), and so inhibit the attachment of HCV onto cells
 CC (especially liver cells), are used to treat HCV infections in mammals,
 CC in particular humans. The specification also describes a method for
 CC identifying a compound which can be used for treating or preventing
 CC HCV in a subject and which can inhibit the attachment of HCV onto cells
 CC by inhibiting the binding of HCV envelope E2 protein to a cellular
 CC protein associated with HCV attachment and entry into cells. The method
 CC comprises incubating the compound, HCV envelope E2 protein or its variant
 CC and a cellular protein capable of specifically binding to the HCV E2
 CC protein under suitable reaction conditions; determining the interactions
 CC between HCV envelope E2 protein and cellular protein in the presence and
 CC absence of the compound; and comparing the interaction to identify a
 CC compound which can inhibit the attachment of HCV onto cells.
 XX
 SQ Sequence 363 AA;
 Query Match 100.0%; Score 20; DB 22; Length 363;
 Best Local Similarity 100.0%; Pred. No. 5.8e-15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CLVDYPYRLWHYPCTINYTI 20
 DB 224 CLVDYPYRLWHYPCTINYTI 243
 RESULT 12
 AAR34438
 ID AAR34438 standard; Protein; 402 AA.
 AC
 XX AAR34438;
 DT 25-MAR-2003 (updated)
 DT 09-AUG-1993 (first entry)
 XX
 DE Sequence of glycoprotein E2/NS1 in clone HCV1.
 XX
 KW Hepatitis C virus; envelope protein; glycoprotein; E2/NS1;
 KW diagnostic reagent.
 XX
 OS Hepatitis C virus.
 XX

PN EP537626-A1.
 XX
 PD 21-APR-1993.
 XX
 PF 08-OCT-1992; 92EP-0117191.
 XX
 PR 08-OCT-1991; 91JP-0260824.
 XX
 XX (NAHE-) NAT INST OF HEALTH.
 XX
 XX Harada S, Honda Y, Miyamura T, Saito I;
 PI
 XX WPI; 1993-127516/16.
 DR
 DR N-PSDB; AAQ40330.
 XX
 XX Diagnostic reagent for hepatitis C virus - comprises second
 PT envelope protein or first non-structural protein encoded by HCV
 PT gene and has sugar chain
 XX
 PS Claim 2; Pages 30-32; 58pp; English.
 XX
 CC Glycoprotein E2/NS1 is derived from the second envelope protein or
 CC first non-structural protein encoded by the genome of HCV. The
 CC nucleic acid is extracted from the serum of the patient of hepatitis
 CC C. The serum is pref. mixed with transfer RNA (tRNA) as a carrier
 CC of virus RNA. As a technique of cloning cDNA from the nucleic acid,
 CC it is preferred to use polymerase chain reaction method. In the
 CC reaction, any commercially available random primers or synthesized
 CC DNA having a base sequence similar to that of primer AS1 may be used
 CC as a primer. Representative examples of sense primers include S1.
 CC (updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 402 AA;
 Query Match 100.0%; Score 20; DB 14; Length 402;
 Best Local Similarity 100.0%; Pred. No. 6.3e-15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CLVDYPYRLWHYPCTINYTI 20
 DB 268 CLVDYPYRLWHYPCTINYTI 287
 RESULT 13
 AAP90183
 ID AAP90183 standard; protein; 454 AA.
 XX
 XX AAP90183;
 DT 25-MAR-2003 (updated)
 DT 01-NOV-1989 (first entry)
 XX
 DE Sequence of hepatitis C virus cDNA insert in clone k9-1.
 XX
 KW Hepatitis C virus; clone k9-1; probe; vaccine.
 XX
 OS Pan troglodytes.
 XX
 XX Key Location/Qualifiers
 FT Region 97..454
 XX
 XX GB2212511-A.
 PN
 XX 26-JUL-1989.
 PD
 XX 18-NOV-1988; 88GB-0027024.
 PF
 XX 18-NOV-1987; 87US-0122714.
 PR 30-DEC-1987; 87US-0139886.
 PR 26-FEB-1988; 88US-0161072.
 PR 26-OCT-1988; 88US-0263584.
 XX
 XX (CHIR) CHIRON CORPORATION.
 PA

XX
PI Houghton M, Choo QL, Kuo G;
XX WPI; 1989-215054/30.
DR N-PSDB; AAN90335.
XX
XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT polypeptide(s) and antibodies for diagnosis, prevention and
PT treatment of infection.
XX
XX Disclosure; fig 46; 235pp; English.
XX
XX The sequence is the peptide encoded by the hepatitis C virus
CC (HCV) cDNA insert in clone K9-1 (see AAN90335). The polypeptides
CC are used to diagnose HCV-induced NANBH, to raise antibodies for
CC immunoassay or treatment, or to produce vaccines.
CC The region shown overlaps the cDNA of AAN90327.
CC (Updated on 25-MAR-2003 to correct PR field.)
XX
XX Sequence 454 AA;
SQ

Query Match 100.0%; Score 20; DB 10; Length 454;
Best Local Similarity 100.0%; Pred. No. 7e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
DB 157 CLVDYPYRLWHYPCTINYTI 176

RESULT 14
AAP92049
ID AAP92049 standard; protein; 454 AA.
XX
AC AAP92049;
XX
DT 25-MAR-2003 (updated)
DT 21-JAN-1991 (first entry)
XX
DE Sequence encoded by segment of the hepatitis C virus (HCV) cDNA sequence
DE in clone K9-1.
XX
XX Non-a non-B hepatitis; probe; vaccine; diagnosis;
KW passive immunotherapy; antigen.
XX
XX Hepatitis C virus.
XX
XX EP318216-A.
XX
PD 31-MAY-1989.
XX
PF 18-NOV-1988; 88EP-0310922.
XX
PR 18-NOV-1987; 87US-0122714.
PR 30-DEC-1987; 87US-0139886.
PR 26-FEB-1988; 88US-0161072.
PR 06-MAY-1988; 88US-0191263.
PR 26-OCT-1988; 88US-0263584.
PR 14-NOV-1988; 88US-0271450.
XX
XX (CHIR) CHIRON CORP.
XX
PI Houghton M, Choo QL, Kuo G;
XX
XX WPI; 1989-159274/22.
DR N-PSDB; AAN92105.
XX
XX Purified hepatitis C virus -
PT and assorted nucleic acids and polypeptide(s)
PT
XX Example; Fig 46-1 - 46-2; 139pp; English.
XX
XX Purified hepatitis C virus (HCV) and purified or recombinant HCV nucleic

XX acids (NAs), encoding HCV polynucleotides or epitopes, and polypeptides
CC are claimed. HCV is a causative agent of non-A, non-B hepatitis (NANBH).
CC The NAs may be used to design probes for detn. of HCV NAs in samples.
CC The polypeptides may be used as immunoassay reagents and vaccines, and
CC to produce antibodies useful for diagnosis and passive immunotherapy.
CC The purified virus may also be used in vaccines.
CC (Updated on 25-MAR-2003 to correct PR field.)
XX
XX Sequence 454 AA;
SQ

Query Match 100.0%; Score 20; DB 10; Length 454;
Best Local Similarity 100.0%; Pred. No. 7e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
DB 157 CLVDYPYRLWHYPCTINYTI 176

RESULT 15
AAB18526
ID AAB18526 standard; Protein; 454 AA.
XX
AC AAB18526;
XX
DT 15-JAN-2001 (first entry)
XX
DE Protein encoded by a novel hepatitis C virus cDNA clone k9-1.
XX
XX Hepatitis C virus; HCV; antisense polynucleotide; polypeptide;
KW viral infectivity; viral replication.
XX
XX Hepatitis C virus.
XX
XX EP1034785-A2.
XX
PD 13-SEP-2000.
XX
PF 16-MAR-1990; 2000EP-0109602.
XX
PR 17-MAR-1989; 89US-0325338.
PR 20-APR-1989; 89US-0341334.
PR 18-MAY-1989; 89US-0355002.
PR 16-MAR-1990; 90EP-0302866.
XX
XX (CHIR) CHIRON CORP.
XX
PI Houghton M, Choo Q, Kuo G;
XX
DR WPI; 2000-566891/53.
DR N-PSDB; AAB75282.
XX
XX Novel composition comprising a hepatitis C virus antisense
PT polynucleotide which is complementary to or corresponds to a sense
PT strand of the virus genome, and selectively hybridises to it -
XX
XX Example; Fig 2; 75pp; English.
XX
XX The specification describes a pharmaceutical composition which
CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
CC HCV is characterized by a positive stranded RNA genome which has
CC 40% homology at the polypeptide level to a HCV polypeptide. The
CC antisense polynucleotide binds to cellular polynucleotides which
CC enhance and/or are required for viral infectivity, replicative
CC ability or chronicity. The antisense polynucleotides may also be
CC designed to bind with high specificity, to be of increased stability,
CC to be stable and to have low toxicity. The composition also comprises
CC an agent which causes viral RNA to be inactive. The composition
CC is used for preventing HCV replication in a system. The present
CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
XX course of the invention.

SQ Sequence 454 AA;

Query Match 100.0%; Score 20; DB 21; Length 454;
Best Local Similarity 100.0%; Pred. No. 7e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYFYLWHYFCTINYTI 20
|||
Db 157 CLVDYFYLWHYFCTINYTI 176

Search completed: November 21, 2003, 20:58:06
Job time : 32.15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.6 Seconds
(without alignments)
79.832 Million cell updates/sec

Title: US-09-973-025-88

Perfect score: 20

Sequence: 1 CLVDYPYRLWHYPCTINYTI 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A COMB.pcp:*
2: /cgn2_6/prodata/1/iaa/5B COMB.pcp:*
3: /cgn2_6/prodata/1/iaa/6A COMB.pcp:*
4: /cgn2_6/prodata/1/iaa/6B COMB.pcp:*
5: /cgn2_6/prodata/1/iaa/PCTUS COMB.pcp:*
6: /cgn2_6/prodata/1/iaa/backfiles1.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-08-612-973-88
2	20	100.0	20	3	US-08-927-597-88
3	20	100.0	20	4	US-08-635-886C-44
4	20	100.0	21	1	US-07-748-292-2
5	20	100.0	21	5	PCT-US92-07813-19
6	20	100.0	68	4	US-08-635-886C-250
7	20	100.0	68	4	US-08-635-886C-252
8	20	100.0	88	1	US-08-440-103-26
9	20	100.0	88	1	US-08-440-103-27
10	20	100.0	88	1	US-08-440-542-26
11	20	100.0	88	1	US-08-440-542-27
12	20	100.0	88	1	US-08-231-368-26
13	20	100.0	88	1	US-08-231-368-27
14	20	100.0	88	1	US-08-440-210-26
15	20	100.0	88	1	US-08-440-210-27
16	20	100.0	88	4	US-09-046-604-26
17	20	100.0	88	4	US-09-046-604-27
18	20	100.0	121	1	US-07-748-292-1
19	20	100.0	179	3	US-08-444-818-77
20	20	100.0	402	1	US-08-460-806-13
21	20	100.0	402	1	US-08-325-630-13
22	20	100.0	403	2	US-08-483-695-39
23	20	100.0	403	2	US-07-965-285-39
24	20	100.0	403	2	US-08-487-231-39
25	20	100.0	403	3	US-09-201-912-39
26	20	100.0	414	1	US-07-748-292-8
27	20	100.0	454	3	US-08-444-818-73

28 20 100.0 480 1 US-08-440-103-14 Sequence 14, Appl
29 20 100.0 480 1 US-08-440-542-14 Sequence 14, Appl
30 20 100.0 480 1 US-08-231-368-14 Sequence 14, Appl
31 20 100.0 480 1 US-08-440-210-14 Sequence 14, Appl
32 20 100.0 480 4 US-09-046-604-14 Sequence 14, Appl
33 20 100.0 621 1 US-07-748-292-7 Sequence 7, Appl
34 20 100.0 622 3 US-08-867-611-34 Sequence 34, Appl
35 20 100.0 622 5 PCT-US92-06965A-4 Sequence 4, Appl
36 20 100.0 663 3 US-08-824-057-3 Sequence 3, Appl
37 20 100.0 663 4 US-09-415-582-3 Sequence 3, Appl
38 20 100.0 663 4 US-09-693-596-4 Sequence 4, Appl
39 20 100.0 738 3 US-08-867-611-35 Sequence 35, Appl
40 20 100.0 738 5 PCT-US92-06965A-5 Sequence 5, Appl
41 20 100.0 2436 3 US-08-444-818-75 Sequence 75, Appl
42 20 100.0 2772 3 US-08-444-818-89 Sequence 89, Appl
43 20 100.0 2894 2 US-08-466-975A-23 Sequence 23, Appl
44 20 100.0 2894 2 US-08-391-671A-23 Sequence 23, Appl
45 20 100.0 2894 3 US-08-467-902A-23 Sequence 23, Appl

ALIGNMENTS

RESULT 1
US-08-612-973-88
; Sequence 88, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612.973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-612-973-88

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
|||||

Db 1 CLVDYPYRLWHYPCTINYTI 20

RESULT 2

US-08-927-597-88
; Sequence 88, Application US/08927597
; Patent No. 6245503
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-927-597-88

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 CLVDYPYRLWHYPCTINYTI 20

RESULT 3

US-08-635-886C-44
; Sequence 44, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555.

; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 44
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-44

Query Match 100.0%; Score 20; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 CLVDYPYRLWHYPCTINYTI 20

RESULT 4

US-07-748-292-2
; Sequence 2, Application US/07748292
; Patent No. 5308750
; GENERAL INFORMATION:
; APPLICANT: MEHTA, SMRITI U.
; APPLICANT: JOHNSON, JILL E.
; APPLICANT: DAILEY, STEPHEN H.
; APPLICANT: DESAI, SURESH M.
; APPLICANT: DEWARE, SUSHIL G.
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES TO PUTATIVE HCV
; TITLE OF INVENTION: E2/NS1 PROTEINS AND METHODS FOR USING SAME
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D-377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/748,292
; FILING DATE: 19910821
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/456,162
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/610,180
; FILING DATE: 07-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4767.US.P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-748-292-2

Query Match 100.0%; Score 20; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 2e-16;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYFYLRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 CLVDYFYLRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |

RESULT 5

PCT-US92-07813-19
; Sequence 19, Application PC/TUS9207813
; GENERAL INFORMATION:
; APPLICANT: LESNIEWSKI, RICHARD R.
; APPLICANT: LEUNG, TAT K.
; TITLE OF INVENTION: HEPATITIS C ASSAY
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES CHAD377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: U.S.A.
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/07813
; FILING DATE: 19920916
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKIP, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4767.P3.03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US92-07813-19

Query Match 100.0%; Score 20; DB 5; Length 21;
Best Local Similarity 100.0%; Pred. No. 2e-16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYFYLRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 CLVDYFYLRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |

RESULT 6

US-08-635-886C-250
; Sequence 250, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; PRIOR FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04

; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 250
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-250

Query Match 100.0%; Score 20; DB 4; Length 68;
Best Local Similarity 100.0%; Pred. No. 5.9e-16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYFYLRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |
Db 37 CLVDYFYLRLWHYPCTINYTI 56
| | | | | | | | | | | | | | | | | | | | | |

RESULT 7

US-08-635-886C-252
; Sequence 252, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert.
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 252
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-252

Query Match 100.0%; Score 20; DB 4; Length 68;
Best Local Similarity 100.0%; Pred. No. 5.9e-16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1-CLVDYFYLRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |
Db 37 CLVDYFYLRLWHYPCTINYTI 56
| | | | | | | | | | | | | | | | | | | | | |

RESULT 8

US-08-440-103-26
; Sequence 26, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/440,103
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-103-26

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 7.4e-16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |
Db 54 CLVDYPYRLWHYPCTINYTI 73

RESULT 9
US-08-440-103-27
; Sequence 27, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,103
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-103-27

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 7.4e-16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |
Db 54 CLVDYPYRLWHYPCTINYTI 73

RESULT 10
US-08-440-542-26
; Sequence 26, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,542
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-542-26

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 7.4e-16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |
Db 54 CLVDYPYRLWHYPCTINYTI 73

RESULT 11
US-08-440-542-27
; Sequence 27, Application US/08440542

Patent No. 5670153
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,542
FILING DATE: 12-MAY-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE: US/08/231,368
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-542-27

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 7.4e-16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPLRLWHYPCTINYTI 20
|||||
Db 54 CLVDYPLRLWHYPCTINYTI 73

RESULT 12
US-08-231-368-26
Sequence 26, Application US/08231368
Patent No. 5756312
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-231-368-26

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 7.4e-16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPLRLWHYPCTINYTI 20
|||||
Db 54 CLVDYPLRLWHYPCTINYTI 73

RESULT 13
US-08-231-368-27
Sequence 27, Application US/08231368
Patent No. 5756312
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-231-368-27

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 7.4e-16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
DB 54 CLVDYPYRLWHYPCTINYTI 73

RESULT 14

US-08-440-210-26
Sequence 26, Application US/08440210
Patent No. 5766845
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,210
FILING DATE: 12-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-210-26

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 7.4e-16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
DB 54 CLVDYPYRLWHYPCTINYTI 73

RESULT 15

US-08-440-210-27
Sequence 27, Application US/08440210
Patent No. 5766845
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.

APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,210
FILING DATE: 12-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-210-27

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 7.4e-16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
DB 54 CLVDYPYRLWHYPCTINYTI 73

Search completed: November 21, 2003, 21:15:17
Job time : 10.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time 19.5 Seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-88

Perfect score: 20

Sequence: 1 CLVDYPLMLHYPCTINVTI 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/prodata/1/pubpaa/ECT_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	10	US-09-973-025-88
2	20	100.0	20	11	US-09-899-303-88
3	20	100.0	20	11	US-09-995-808-88
4	20	100.0	20	11	US-09-995-860-88
5	20	100.0	20	12	US-09-995-791-88
6	20	100.0	254	10	US-09-407-430-3
7	20	100.0	363	10	US-09-407-430-2
8	20	100.0	637	12	US-10-187-257-4
9	20	100.0	637	12	US-10-265-083-2
10	20	100.0	2894	10	US-09-941-611-23
11	20	100.0	2894	15	US-10-044-995-23
12	20	100.0	3011	9	US-09-916-359-2
13	20	100.0	3011	16	US-10-232-643-6
14	18	90.0	176	10	US-09-921-397-81
15	18	90.0	250	10	US-09-952-572-8

16	18	90.0	350	10	US-09-929-955-4	Sequence 4, Appli
17	18	90.0	350	14	US-10-104-966-4	Sequence 4, Appli
18	18	90.0	363	12	US-10-128-587A-97	Sequence 97, Appl
19	18	90.0	363	15	US-10-128-590-97	Sequence 97, Appl
20	18	90.0	3011	9	US-09-742-659-4	Sequence 4, Appli
21	18	90.0	3011	10	US-09-238-076-20	Sequence 20, Appl
22	18	90.0	3011	10	US-09-952-572-9	Sequence 9, Appli
23	18	90.0	3011	10	US-09-929-955-1	Sequence 1, Appli
24	18	90.0	3011	10	US-09-747-419-20	Sequence 20, Appl
25	18	90.0	3011	11	US-09-891-894-3	Sequence 3, Appli
26	18	90.0	3011	11	US-09-995-937-20	Sequence 20, Appl
27	18	90.0	3011	11	US-09-917-563-20	Sequence 20, Appl
28	18	90.0	3011	12	US-10-184-150-3	Sequence 3, Appli
29	18	90.0	3011	14	US-10-104-966-1	Sequence 1, Appli
30	18	90.0	3011	15	US-10-259-275-20	Sequence 20, Appl
31	18	90.0	3012	10	US-09-238-076-2	Sequence 2, Appli
32	18	90.0	3012	11	US-09-995-937-2	Sequence 2, Appli
33	18	90.0	3012	11	US-09-917-563-2	Sequence 2, Appli
34	13	65.0	347	11	US-09-194-949-9	Sequence 9, Appli
35	13	65.0	539	11	US-09-194-949-11	Sequence 11, Appl
36	13	65.0	2985	15	US-10-259-275-40	Sequence 40, Appl
37	12	60.0	2940	12	US-10-226-629A-13	Sequence 13, Appl
38	11	55.0	45	16	US-10-318-200-16	Sequence 16, Appl
39	11	55.0	290	12	US-10-128-587A-3	Sequence 3, Appli
40	11	55.0	290	15	US-10-128-590-3	Sequence 3, Appli
41	11	55.0	301	12	US-10-128-587A-5	Sequence 5, Appli
42	11	55.0	301	15	US-10-128-590-5	Sequence 5, Appli
43	11	55.0	314	10	US-09-973-025-42	Sequence 42, Appl
44	11	55.0	314	11	US-09-899-303-42	Sequence 42, Appl
45	11	55.0	314	11	US-09-995-808-42	Sequence 42, Appl

ALIGNMENTS

RESULT 1

US-09-973-025-88
; Sequence 88, Application US/09973025
; Publication No. US20020182706A1

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

BOSMAN, FONS

DE MARTYNOFF, GUY

BUYSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/973,025

FILING DATE: 10-Oct-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 88:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 88:
US-09-973-025-88

Query Match 100.0%; Score 20; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.3e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
|||
DB 1 CLVDYPYRLWHYPCTINYTI 20
|||

RESULT 2
US-09-899-303-88
; Sequence 88, Application US/09899303
; Publication No. US20030036110A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 88:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 88:
US-09-899-303-88

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.3e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
|||

Db 1 CLVDYPYRLWHYPCTINYTI 20

RESULT 3
US-09-995-808-88
; Sequence 88, Application US/09995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent in 3.1
; SEQ ID NO 88
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-88

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.3e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
|||
DB 1 CLVDYPYRLWHYPCTINYTI 20
|||

RESULT 4
US-09-995-860-88
; Sequence 88, Application US/09995860
; Publication No. US20030118603A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent in 3.1
; SEQ ID NO 88
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-88

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.3e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
|||
DB 1 CLVDYPYRLWHYPCTINYTI 20
|||

RESULT 5
US-09-995-791-88
; Sequence 88, Application US/09995791
; Publication No. US20030147918A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent in 3.1
; SEQ ID NO 88

```
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-88

Query Match      100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.3e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLVDYPYRLWHYPCTINYTI 20
   |||||
Db 1 CLVDYPYRLWHYPCTINYTI 20
   |||||

RESULT 6
US-09-407-430-3
; Sequence 3, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-407-430-3

Query Match      100.0%; Score 20; DB 10; Length 254;
Best Local Similarity 100.0%; Pred. No. 5.6e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLVDYPYRLWHYPCTINYTI 20
   |||||
Db 202 CLVDYPYRLWHYPCTINYTI 221
   |||||

RESULT 7
US-09-407-430-2
; Sequence 2, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-09-407-430-2

Query Match      100.0%; Score 20; DB 10; Length 363;
Best Local Similarity 100.0%; Pred. No. 7.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLVDYPYRLWHYPCTINYTI 20
   |||||
Db 224 CLVDYPYRLWHYPCTINYTI 243
   |||||

RESULT 8

US-10-187-257-4
; Sequence 4, Application US/10187257
; Publication No. US20030138458A1
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: CORTES, Steve
; APPLICANT: O'HAGAN, Derek
; TITLE OF INVENTION: HCV E1E2 VACCINE COMPOSITIONS
; FILE REFERENCE: 2302-17206
; CURRENT APPLICATION NUMBER: US/10/187,257
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
US-10-187-257-4

Query Match      100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLVDYPYRLWHYPCTINYTI 20
   |||||
Db 435 CLVDYPYRLWHYPCTINYTI 454
   |||||

RESULT 9
US-10-265-083-2
; Sequence 2, Application US/10265083
; Publication No. US20030170273A1
; GENERAL INFORMATION:
; APPLICANT: O'HAGAN, Derek
; APPLICANT: VALIANTE, Nicholas
; TITLE OF INVENTION: ADJUVANT COMPOSITIONS
; FILE REFERENCE: 2302-18357.30
; CURRENT APPLICATION NUMBER: US/10/265,083
; CURRENT FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7
US-10-265-083-2

Query Match      100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLVDYPYRLWHYPCTINYTI 20
   |||||
Db 435 CLVDYPYRLWHYPCTINYTI 454
   |||||

RESULT 10
US-09-941-611-23
; Sequence 23, Application US/09941611
; Patent No. US20020106640A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSWIN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; ANTIBODIES TO HEPATITIS C VIRUS
```

```

;
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
;   ADDRESS: NIXON & VANDERHYE P.C.
;   STREET: 1100 NORTH GLEBE ROAD
;   CITY: ARLINGTON
;   STATE: VA
;   COUNTRY: USA
;   ZIP: 22201
;
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/941,611
;   FILING DATE: 30-Aug-2001
;   CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/391,671
;   FILING DATE: 1995-02-21
;   APPLICATION NUMBER: WO PCT/EP91/02409
;   FILING DATE: 13-DEC-1991
;   APPLICATION NUMBER: EP 90124241.2
;   FILING DATE: 14-DEC-1990
;
; ATTORNEY/AGENT INFORMATION:
;   NAME: SADOFF, B.J.
;   REGISTRATION NUMBER: 36,663
;   REFERENCE/DOCKET NUMBER: 1487-5
;
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 7038164000
;   TELEFAX: 7038164100
;
; INFORMATION FOR SEQ ID NO: 23:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 2894 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   SEQUENCE DESCRIPTION: SEQ ID NO: 23:
;
US-09-941-611-23

Query Match          100.0%; Score 20; DB 10; Length 2894;
Best Local Similarity 100.0%; Pred. No. 4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CLVDYPYRLWHYPCTINYTI 20
      ||||||||||||||||||
Db      607 CLVDYPYRLWHYPCTINYTI 626

RESULT 11
US-10-044-995-23
; Sequence 23, Application US/10044995
; Publication No. US20030049685A1
; GENERAL INFORMATION:
; APPLICANT: POLEY, ROBERT J
;   POLEY, DIRK
;   MAERTENS, GEERT
;   VAN HEUVERSMUN, HUGO
;
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
;   ANTIBODIES TO HEPATITIS C VIRUS
;
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: NIXON & VANDERHYE P.C.
;   STREET: 1100 NORTH GLEBE ROAD
;   CITY: ARLINGTON
;   STATE: VA
;   COUNTRY: USA
;   ZIP: 22201
;
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk

```

```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/044,995
;   FILING DATE: 15-Jan-2002
;   CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/391,671
;   FILING DATE: <Unknown>
;   APPLICATION NUMBER: US 07/920,286
;   FILING DATE: 14-OCT-1992
;   APPLICATION NUMBER: WO PCT/EP91/02409
;   FILING DATE: 13-DEC-1991
;   APPLICATION NUMBER: EP 90124241.2
;   FILING DATE: 14-DEC-1990
;
; ATTORNEY/AGENT INFORMATION:
;   NAME: SADOFF, B.J.
;   REGISTRATION NUMBER: 36,663
;   REFERENCE/DOCKET NUMBER: 1487-5
;
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 7038164000
;   TELEFAX: 7038164100
;
; INFORMATION FOR SEQ ID NO: 23:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 2894 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   SEQUENCE DESCRIPTION: SEQ ID NO: 23:
;
US-10-044-995-23

Query Match          100.0%; Score 20; DB 15; Length 2894;
Best Local Similarity 100.0%; Pred. No. 4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CLVDYPYRLWHYPCTINYTI 20
      ||||||||||||||||||
Db      607 CLVDYPYRLWHYPCTINYTI 626

RESULT 12
US-09-916-359-2
; Sequence 2, Application US/09916359
; Patent No. US20020034734A1
; GENERAL INFORMATION:
; APPLICANT: Veronique Barban
;
; TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
;   TREATING C HEPATITIS
;
; FILE REFERENCE: PMCF97-03A
;
; CURRENT APPLICATION NUMBER: US/09/916,359
;
; CURRENT FILING DATE: 2001-07-26
;
; PRIOR APPLICATION NUMBER: 09/388,874
;
; PRIOR FILING DATE: 1999-09-02
;
; PRIOR APPLICATION NUMBER: 97/02,887
;
; PRIOR FILING DATE: 1997-03-06
;
; NUMBER OF SEQ ID NOS: 2
;
; SOFTWARE: FastSeq for Windows Version 3.0
;
; SEQ ID NO 2
;   LENGTH: 3011
;   TYPE: PRT
;   ORGANISM: Virus
;
US-09-916-359-2

Query Match          100.0%; Score 20; DB 9; Length 3011;
Best Local Similarity 100.0%; Pred. No. 4.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CLVDYPYRLWHYPCTINYTI 20
      ||||||||||||||||||

```

Db 607 CLVDYPRLWHYPCTINYTI 626

RESULT 13
US-10-232-643-6
; Sequence 6, Application US/10232643
; Publication No. US20030129586A1
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, MICHAEL
; CHOO, QUI-LIM
; HAN, JANG
; CHOE, JOONHO
; TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
; HELICASE ACTIVITY AND IMPROVED SOLUBILITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: Intellectual Property - R440, P.O. Box 8097
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/232,643
; FILING DATE: 30-AUG-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/483,799
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/529,169
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0100.005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-3274
; TELEFAX: (510) 655-3542
; TELEX: n/a
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 9
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Lys or Arg"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 11
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Asn or Thr"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 176
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Ile or Thr"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 334
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Met or Val"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2502
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Leu or Phe"
; LOCATION: 603
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Leu or Ile"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 849
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Tyr or Asn"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1114
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Pro or Ser"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1117
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Ser or Thr"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1276
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Pro or Leu"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1454
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Cys or Tyr"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1471
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Thr or Ser"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1877
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Glu or Gly"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1948
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Leu or His"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1949
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Ser or Cys"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2021
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Gly or Val"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2349
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Thr or Ser"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2385
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Tyr or Phe"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2386
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Ser or Ala"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2502
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Leu or Phe"

```
;
;
; NAME/KEY: Duplication
; LOCATION: 2690
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Arg or Gly"
;
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2921
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Arg or Gly"
;
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2996
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Leu or Pro"
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-232-643-6

Query Match      100.0%; Score 20; DB 16; Length 3011;
Best Local Similarity 100.0%; Pred. No. 4.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 CLVDYPYRLWHYPCTINYTI 20
Db      607 CLVDYPYRLWHYPCTINYTI 626

RESULT 14
US-09-921-397-81
; Sequence 81, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 81
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-81

Query Match      90.0%; Score 18; DB 10; Length 176;
Best Local Similarity 100.0%; Pred. No. 5.6e-12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      3 VDYPYRLWHYPCTINYTI 20
Db      141 VDYPYRLWHYPCTINYTI 158

RESULT 15
US-09-952-572-8
; Sequence 8, Application US/09952572
; Patent No. US20020119495A1
; GENERAL INFORMATION:
; APPLICANT: HAWAII BIOTECHNOLOGY GROUP, Inc.
; APPLICANT: NAKANO, Eileen
; APPLICANT: CLEMENTS, David
; APPLICANT: HUMPHREYS, Tom
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITION OF HEPATITIS C
; FILE REFERENCE: HAWBIO1100
; CURRENT APPLICATION NUMBER: US/09/952,572
; CURRENT FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: US 60/230,927
; PRIOR FILING DATE: 2000-09-13

;
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-09-952-572-8

Query Match      90.0%; Score 18; DB 10; Length 250;
Best Local Similarity 100.0%; Pred. No. 7.4e-12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      3 VDYPYRLWHYPCTINYTI 20
Db      198 VDYPYRLWHYPCTINYTI 215

Search completed: November 21, 2003, 22:19:40
Job time : 19.5 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 163.1 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-88

Perfect score: 20
Sequence: 1 CLVDIPYRLWHYPTINYTI 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents AA Main:*

1:	/cgn2_6/ptodata/1/paa/PCTUS COMB.pcp.*
2:	/cgn2_6/ptodata/1/paa/US06 COMB.pcp.*
3:	/cgn2_6/ptodata/1/paa/US07 COMB.pcp.*
4:	/cgn2_6/ptodata/1/paa/US08 COMB.pcp.*
5:	/cgn2_6/ptodata/1/paa/US081 COMB.pcp.*
6:	/cgn2_6/ptodata/1/paa/US082 COMB.pcp.*
7:	/cgn2_6/ptodata/1/paa/US083 COMB.pcp.*
8:	/cgn2_6/ptodata/1/paa/US084 COMB.pcp.*
9:	/cgn2_6/ptodata/1/paa/US085 COMB.pcp.*
10:	/cgn2_6/ptodata/1/paa/US086 COMB.pcp.*
11:	/cgn2_6/ptodata/1/paa/US087 COMB.pcp.*
12:	/cgn2_6/ptodata/1/paa/US088 COMB.pcp.*
13:	/cgn2_6/ptodata/1/paa/US089 COMB.pcp.*
14:	/cgn2_6/ptodata/1/paa/US090 COMB.pcp.*
15:	/cgn2_6/ptodata/1/paa/US091 COMB.pcp.*
16:	/cgn2_6/ptodata/1/paa/US092 COMB.pcp.*
17:	/cgn2_6/ptodata/1/paa/US093 COMB.pcp.*
18:	/cgn2_6/ptodata/1/paa/US094 COMB.pcp.*
19:	/cgn2_6/ptodata/1/paa/US095 COMB.pcp.*
20:	/cgn2_6/ptodata/1/paa/US096 COMB.pcp.*
21:	/cgn2_6/ptodata/1/paa/US097A COMB.pcp.*
22:	/cgn2_6/ptodata/1/paa/US097B COMB.pcp.*
23:	/cgn2_6/ptodata/1/paa/US098 COMB.pcp.*
24:	/cgn2_6/ptodata/1/paa/US099A COMB.pcp.*
25:	/cgn2_6/ptodata/1/paa/US099B COMB.pcp.*
26:	/cgn2_6/ptodata/1/paa/US100 COMB.pcp.*
27:	/cgn2_6/ptodata/1/paa/US101 COMB.pcp.*
28:	/cgn2_6/ptodata/1/paa/US102 COMB.pcp.*
29:	/cgn2_6/ptodata/1/paa/US103 COMB.pcp.*
30:	/cgn2_6/ptodata/1/paa/US104 COMB.pcp.*
31:	/cgn2_6/ptodata/1/paa/US106 COMB.pcp.*
32:	/cgn2_6/ptodata/1/paa/US60 COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	DB ID	Description
1	20	100.0	20 13	US-08-974-685-44 Sequence 44, Appl

2	20	100.0	20 13	US-08-974-690-44 Sequence 44, Appl
3	20	100.0	20 13	US-08-974-690A-44 Sequence 44, Appl
4	20	100.0	20 13	US-08-974-690B-44 Sequence 44, Appl
5	20	100.0	20 13	US-08-974-690C-44 Sequence 44, Appl
6	20	100.0	20 23	US-09-899-303-88 Sequence 88, Appl
7	20	100.0	20 23	US-09-899-303A-88 Sequence 88, Appl
8	20	100.0	20 25	US-09-973-025-88 Sequence 88, Appl
9	20	100.0	20 25	US-09-995-791-88 Sequence 88, Appl
10	20	100.0	20 25	US-09-995-808-88 Sequence 88, Appl
11	20	100.0	20 25	US-09-995-860-88 Sequence 88, Appl
12	20	100.0	20 26	US-10-020-510-88 Sequence 88, Appl
13	20	100.0	20 29	US-10-321-798-88 Sequence 88, Appl
14	20	100.0	21 9	US-08-507-740-19 Sequence 19, Appl
15	20	100.0	21 13	US-08-905-054-19 Sequence 19, Appl
16	20	100.0	21 13	US-08-905-054B-19 Sequence 19, Appl
17	20	100.0	68 13	US-08-974-690C-250 Sequence 250, App
18	20	100.0	68 13	US-08-974-690C-252 Sequence 252, App
19	20	100.0	88 8	US-08-471-498-26 Sequence 26, Appl
20	20	100.0	88 8	US-08-471-498-27 Sequence 27, Appl
21	20	100.0	112 8	US-08-436-966-19 Sequence 19, Appl
22	20	100.0	179 8	US-08-403-590B-77 Sequence 77, Appl
23	20	100.0	179 8	US-08-444-112-77 Sequence 77, Appl
24	20	100.0	254 1	PCT-US00-26395-3 Sequence 3, Appli
25	20	100.0	254 18	US-09-407-430-3 Sequence 3, Appli
26	20	100.0	333 30	US-10-445-724-3 Sequence 3, Appli
27	20	100.0	333 32	US-60-409-909-4 Sequence 4, Appli
28	20	100.0	363 1	PCT-US00-26395-2 Sequence 2, Appli
29	20	100.0	363 18	US-09-407-430-2 Sequence 2, Appli
30	20	100.0	414 3	US-07-748-561-2 Sequence 2, Appli
31	20	100.0	454 8	US-08-403-590B-73 Sequence 73, Appl
32	20	100.0	454 8	US-08-444-112-73 Sequence 73, Appl
33	20	100.0	480 8	US-08-471-498-14 Sequence 14, Appl
34	20	100.0	531 21	US-09-721-480-7 Sequence 7, Appli
35	20	100.0	622 3	US-07-748-561-4 Sequence 4, Appli
36	20	100.0	622 3	US-07-989-843-34 Sequence 34, Appl
37	20	100.0	622 8	US-08-463-843-34 Sequence 34, Appl
38	20	100.0	622 8	US-08-463-884-34 Sequence 34, Appl
39	20	100.0	622 20	US-09-690-359-34 Sequence 34, Appl
40	20	100.0	637 1	PCT-US02-20676-4 Sequence 4, Appli
41	20	100.0	637 1	PCT-US02-30423-2 Sequence 2, Appli
42	20	100.0	637 1	PCT-US02-31486-2 Sequence 2, Appli
43	20	100.0	637 27	US-10-187-257-4 Sequence 4, Appli
44	20	100.0	637 28	US-10-285-083-2 Sequence 2, Appli
45	20	100.0	663 6	US-08-282-959-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-08-974-685-44
; Sequence 44, Application US/08974685
; GENERAL INFORMATION:

APPLICANT: LEROUX-ROELS, GEERT
DELEYS, ROBERT
MAERTENS, GEERT

TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 181

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHUYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25 (BPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,685
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 44:

US-08-974-685-44

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPRRLWHYPCTINYTI 20
|||||
DB 1 CLVDYPRRLWHYPCTINYTI 20

RESULT 2

US-08-974-690-44

Sequence 44, Application US/08974690
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
APPLICANT: MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 166
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (BPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/635,886.
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4100
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

US-08-974-690-44

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPRRLWHYPCTINYTI 20
|||||
DB 1 CLVDYPRRLWHYPCTINYTI 20

RESULT 3

US-08-974-690A-44

Sequence 44, Application US/08974690A
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
APPLICANT: MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (BPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690A
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 44:

US-08-974-690A-44

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPRRLWHYPCTINYTI 20
|||||
DB 1 CLVDYPRRLWHYPCTINYTI 20

RESULT 4

US-08-974-690B-44

Sequence 44, Application US/08974690B
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
APPLICANT: MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303A
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 88:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 88:
US-09-899-303A-88

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |
DB 1 CLVDYPYRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |

RESULT 8
US-09-973-025-88
; Sequence 88, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 88:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 88:
US-09-973-025-88

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |
DB 1 CLVDYPYRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |

RESULT 9
US-09-995-791-88
; Sequence 88, Application US/09995791
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 88
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-88

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |
DB 1 CLVDYPYRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |

RESULT 10
US-09-995-808-88
; Sequence 88, Application US/09995808
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In 3.1
; SEQ ID NO 88
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-88

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
|||||
Db 1 CLVDYPYRLWHYPCTINYTI 20
|||||

RESULT 11
US-09-995-860-88
; Sequence 88, Application US/0995860
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 88
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-88

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
|||||
Db 1 CLVDYPYRLWHYPCTINYTI 20
|||||

RESULT 12
US-10-020-510-88
; Sequence 88, Application US/10020510
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-72
; CURRENT APPLICATION NUMBER: US/10/020,510
; CURRENT FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 88
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-020-510-88

Query Match 100.0%; Score 20; DB 26; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
|||||
Db 1 CLVDYPYRLWHYPCTINYTI 20
|||||

RESULT 13
US-10-321-798-88
; Sequence 88, Application US/10321798
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-93
; CURRENT APPLICATION NUMBER: US/10/321,798
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: 60/418,358
; PRIOR FILING DATE: 2002-10-16

; PRIOR APPLICATION NUMBER: 10/020,510
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 88
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-321-798-88

Query Match 100.0%; Score 20; DB 29; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
|||||
Db 1 CLVDYPYRLWHYPCTINYTI 20
|||||

RESULT 14
US-08-507-740-19
; Sequence 19, Application US/08507740
; GENERAL INFORMATION:
; APPLICANT: LESNIEWSKI, RICHARD R.
; APPLICANT: LEUNG, TAT K.
; TITLE OF INVENTION: HEPATITIS C ASSAY
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES CHAD377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: U.S.A.
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/507,740
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/373,920
; FILING DATE:
; APPLICATION NUMBER: US/08/183,207
; FILING DATE:
; APPLICATION NUMBER: US 07/760,292
; FILING DATE: 16-SEP-1991
; APPLICATION NUMBER: US 07/456,162
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/610,180
; FILING DATE: 07-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKIP, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4767.US.P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-507-740-19

Query Match 100.0%; Score 20; DB 9; Length 21;
Best Local Similarity 100.0%; Pred. No. 3e-15;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLVDYPYRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 CLVDYPYRLWHYPCTINYTI 20

RESULT 15
US-08-905-054-19
; Sequence 19, Application US/08905054
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Lesniewski, Richard R.
; APPLICANT: Leung, Tat K.
; TITLE OF INVENTION: HEPATITIS C ASSAY
; FILE REFERENCE: 4767.US.C7
; CURRENT APPLICATION NUMBER: US/08/905,054
; CURRENT FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/707,355
; PRIOR FILING DATE: 1996-09-04
; PRIOR APPLICATION NUMBER: US 08/507,740
; PRIOR FILING DATE: 1995-07-26
; PRIOR APPLICATION NUMBER: US 08/373,920
; PRIOR FILING DATE: 1995-01-17
; PRIOR APPLICATION NUMBER: US 08/183,207
; PRIOR FILING DATE: 1994-01-18
; PRIOR APPLICATION NUMBER: US 07/760,292
; PRIOR FILING DATE: 1991-09-16
; PRIOR APPLICATION NUMBER: US 07/610,180
; PRIOR FILING DATE: 1990-11-07
; PRIOR APPLICATION NUMBER: US 07/456,162
; PRIOR FILING DATE: 1989-12-22
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-08-905-054-19

Query Match 100.0%; Score 20; DB 13; Length 21;
Best Local Similarity 100.0%; Pred. No. 3e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLVDYPYRLWHYPCTINYTI 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 CLVDYPYRLWHYPCTINYTI 20

Search completed: November 21, 2003, 22:09:53
Job time : 163.1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.55 Seconds
(without alignments)
117.536 Million cell updates/sec

Title: us-09-973-025-88

Perfect score: 20
Sequence: 1 CLVDYPYRLWHYPCTINYTI 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA_New.*
1: /cgn2_6/prodata/1/paa/PCT_NEW_COMB.pcp.*
2: /cgn2_6/prodata/1/paa/US06_NEW_COMB.pcp.*
3: /cgn2_6/prodata/1/paa/US07_NEW_COMB.pcp.*
4: /cgn2_6/prodata/1/paa/US08_NEW_COMB.pcp.*
5: /cgn2_6/prodata/1/paa/US09_NEW_COMB.pcp.*
6: /cgn2_6/prodata/1/paa/US10_NEW_COMB.pcp.*
7: /cgn2_6/prodata/1/paa/US60_NEW_COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	6	US-10-651-165-44
2	20	100.0	30	6	US-10-296-734-488
3	20	100.0	68	6	US-10-651-165-250
4	20	100.0	68	6	US-10-651-165-252
5	20	100.0	333	1	PCT-US03-19834-3
6	20	100.0	333	6	US-10-651-165-262
7	20	100.0	637	1	PCT-US03-33610-4
8	20	100.0	1997	6	US-10-296-734-816
9	20	100.0	3011	1	PCT-US03-19834-2
10	20	100.0	3011	6	US-10-296-734-406
11	20	100.0	5985	6	US-10-296-734-810
12	18	90.0	68	6	US-10-651-165-251
13	15	75.0	68	6	US-10-651-165-254
14	15	75.0	68	6	US-10-651-165-255
15	15	75.0	68	6	US-10-651-165-256
16	15	75.0	68	6	US-10-651-165-262
17	15	75.0	70	6	US-10-651-165-264
18	13	65.0	30	6	US-10-296-734-490
19	13	65.0	30	6	US-10-651-165-253
20	13	65.0	68	6	US-10-651-165-257
21	13	65.0	68	6	US-10-651-165-258
22	13	65.0	68	6	US-10-651-165-259
23	13	65.0	68	6	US-10-651-165-260
24	13	65.0	68	6	US-10-651-165-261
25	13	65.0	347	6	US-10-664-391-9
26	13	65.0	539	6	US-10-664-391-11

Sequence 3, Appli
Sequence 814, App
Sequence 211, App
Sequence 133, App
Sequence 108, App
Sequence 263, App
Sequence 5, Appli
Sequence 29, Appli
Sequence 28, Appli
Sequence 27, Appli
Sequence 59, Appli
Sequence 61, Appli
Sequence 93, Appli
Sequence 42, Appli
Sequence 124, App
Sequence 43, Appli
Sequence 45, Appli
Sequence 172, App
Sequence 123, App

13 65.0 1026 1 PCT-US03-20409-3
13 65.0 2010 6 US-10-296-734-814
29 13 65.0 2280 1 PCT-US03-20322-211
30 12 60.0 20 6 US-10-651-165-133
31 12 60.0 70 6 US-10-651-165-108
32 12 60.0 70 6 US-10-651-165-263
33 12 60.0 3033 6 US-10-009-002-5
34 11 55.0 25 6 US-10-685-435-29
35 11 55.0 33 6 US-10-685-435-28
36 11 55.0 45 6 US-10-685-435-27
37 10 50.0 10 6 US-10-440-390-59
38 10 50.0 10 6 US-10-440-390-61
39 10 50.0 10 6 US-10-440-390-93
40 9 45.0 9 6 US-10-440-390-42
41 8 40.0 8 6 US-10-651-165-124
42 8 40.0 20 6 US-10-651-165-43
43 8 40.0 20 6 US-10-651-165-45
44 7 35.0 7 6 US-10-651-165-172
45 7 35.0 8 6 US-10-651-165-123

ALIGNMENTS

RESULT 1
US-10-651-165-44
; Sequence 44, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-44

Query Match 100.0%; Score 20; DB 6; Length 20;
Best Local Similarity 100.0%; Pred No. 1.4e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
Db 1 CLVDYPYRLWHYPCTINYTI 20

RESULT 2
US-10-296-734-488
; Sequence 488, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2

```
; SEQ ID NO 488
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 41
US-10-296-734-488

Query Match          100.0%; Score 20; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.9e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
Db 9 CLVDYPYRLWHYPCTINYTI 28

RESULT 3
US-10-651-165-250
; Sequence 250, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651.165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974.690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 250
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-250

Query Match          100.0%; Score 20; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 3.3e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
Db 37 CLVDYPYRLWHYPCTINYTI 56

RESULT 4
US-10-651-165-252
; Sequence 252, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651.165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974.690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 252
```

```
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-252

Query Match          100.0%; Score 20; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 3.3e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
Db 37 CLVDYPYRLWHYPCTINYTI 56

RESULT 5
PCT-US03-19834-3
; Sequence 3, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; FILE REFERENCE: IOWA:045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-3

Query Match          100.0%; Score 20; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
Db 225 CLVDYPYRLWHYPCTINYTI 244

RESULT 6
US-10-655-562-4
; Sequence 4, Application US/10655562
; GENERAL INFORMATION:
; APPLICANT: BISHOP-HURLEY, SHARON L.
; APPLICANT: SCHMIDT, FRANCIS J.
; APPLICANT: SMITH, ARNOLD L.
; TITLE OF INVENTION: PHAGE-DISPLAY PEPTIDES AS NOVEL ANTIMICROBIAL AGENTS
; FILE REFERENCE: UWMO:022US
; CURRENT APPLICATION NUMBER: US/10/655,562
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: 60,409,909
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-655-562-4

Query Match          100.0%; Score 20; DB 6; Length 333;
Best Local Similarity 100.0%; Pred. No. 1e-14;
```



```
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
Db 225 CLVDYPYRLWHYPCTINYTI 244
|||||

RESULT 7
PCT-US03-33610-4
; Sequence 4, Application PC/TUS0333610
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: SELBY, Mark
; APPLICANT: PALLIARD, Xavier
; TITLE OF INVENTION: ACTIVATION OF HCV-SPECIFIC T-CELLS
; FILE REFERENCE: 2300-1612.60
; CURRENT APPLICATION NUMBER: PCT/US03/33610
; CURRENT FILING DATE: 2003-10-24
; PRIOR APPLICATION NUMBER: 10/281,341
; PRIOR FILING DATE: 2002-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
PCT-US03-33610-4

Query Match 100.0%; Score 20; DB 1; Length 637;
Best Local Similarity 100.0%; Pred. No. 1.6e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
Db 435 CLVDYPYRLWHYPCTINYTI 454
|||||

RESULT 8
US-10-296-734-816
; Sequence 816, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 816
; LENGTH: 1997
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette C
US-10-296-734-816

Query Match 100.0%; Score 20; DB 6; Length 1997;
Best Local Similarity 100.0%; Pred. No. 3.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
Db 1974 CLVDYPYRLWHYPCTINYTI 1993
|||||

RESULT 9
PCT-US03-19834-2
```

```
; Sequence 2, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; TITLE OF INVENTION: METHODS FOR TREATING HYPERLIPIDEMIA
; FILE REFERENCE: IOWA:045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-2

Query Match 100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 4.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
Db 607 CLVDYPYRLWHYPCTINYTI 626
|||||

RESULT 10
US-10-296-734-406
; Sequence 406, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a consensus polyprotein
US-10-296-734-406

Query Match 100.0%; Score 20; DB 6; Length 3011;
Best Local Similarity 100.0%; Pred. No. 4.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
Db 607 CLVDYPYRLWHYPCTINYTI 626
|||||

RESULT 11
US-10-296-734-810
; Sequence 810, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
```

```
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 810
; LENGTH: 5985
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la savine
US-10-296-734-810
```

```
Query Match          100.0%; Score 20; DB 6; Length 5985;
Best Local Similarity 100.0%; Pred. No. 7.6e-14; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 CLVDYPYRLWHYPCTINYTI 20
   |||||
Db 5964 CLVDYPYRLWHYPCTINYTI 5983
```

```
RESULT 12
US-10-651-165-251
; Sequence 251, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 251
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-251
```

```
Query Match          90.0%; Score 18; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 4.8e-13; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 3 VDYPYRLWHYPCTINYTI 20
   |||||
Db 39 VDYPYRLWHYPCTINYTI 56
```

```
RESULT 13
US-10-651-165-254
; Sequence 254, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 254
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-254
```

```
Query Match          75.0%; Score 15; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 8.2e-10; Indels 0; Gaps 0;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 CLVDYPYRLWHYPCT 15
   |||||
Db 37 CLVDYPYRLWHYPCT 51
```

```
RESULT 14
US-10-651-165-255
; Sequence 255, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 255
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-255
```

```
Query Match          75.0%; Score 15; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 8.2e-10; Indels 0; Gaps 0;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 CLVDYPYRLWHYPCT 15
   |||||
Db 37 CLVDYPYRLWHYPCT 51
```

```
RESULT 15
US-10-651-165-256
; Sequence 256, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 256
; LENGTH: 68
```

; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-256

Query Match 75.0%; Score 15; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 8.2e-10;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHPCT 15
|||
Db 37 CLVDYPYRLWHPCT 51

Search completed: November 21, 2003, 22:12:59
Job time : 9.55 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:46:27 ; Search time 9.25 Seconds
(without alignments)
207.932 Million cell updates/sec

Title: US-09-973-025-88
Perfect score: 20
Sequence: 1 CLVDYPRYLWHYPCTINYTI 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 76:.*
1: Pir1:.*
2: Pir2:.*
3: Pir3:.*
4: Pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	3011	1 GNMWC3	genome polyprotein
2	20	100.0	3011	1 S40770	genome polyprotein
3	18	90.0	3011	1 GNMVCH	genome polyprotein
4	15	75.0	350	2 S35631	genome polyprotein
5	15	75.0	640	2 JQ1584	genome polyprotein
6	15	75.0	3033	1 GNMWJ8	genome polyprotein
7	13	65.0	782	2 S19875	genome polyprotein
8	13	65.0	787	2 PN0677	hypothetical prote
9	13	65.0	3010	1 GNMVTC	genome polyprotein
10	13	65.0	3010	1 GNMWCJ	genome polyprotein
11	13	65.0	3010	1 A45573	genome polyprotein
12	13	65.0	3010	1 S18030	genome polyprotein
13	13	65.0	3010	1 GNMVTV	genome polyprotein
14	12	60.0	3033	1 JQ1303	genome polyprotein
15	11	55.0	716	2 JQ1366	polyprotein - hepa
16	11	55.0	782	2 S19876	genome polyprotein
17	11	55.0	782	2 S18031	genome polyprotein
18	11	55.0	782	2 S18032	genome polyprotein
19	11	55.0	3014	1 JC5620	genome polyprotein
20	6	30.0	68	2 A95993	hypothetical prote
21	6	30.0	265	2 C45392	orf3 protein - por
22	6	30.0	265	2 D36861	orf3 protein - iel
23	6	30.0	312	1 B64170	tRNA-pseudouridine
24	5	25.0	88	2 S36917	ubiquinol-cytochro
25	5	25.0	94	2 I67528	CD33 antigen homol
26	5	25.0	95	2 I67527	CD33 antigen homol
27	5	25.0	116	1 HVMS44	Ig heavy chain pre
28	5	25.0	117	1 G2NSU1	Ig heavy chain v r
29	5	25.0	117	1 G2MS73	Ig heavy chain v r

30	5	25.0	118	2 A47329	Ig heavy chain v r
31	5	25.0	119	1 AVMSX4	Ig heavy chain v r
32	5	25.0	119	1 AVNST6	Ig heavy chain v r
33	5	25.0	119	1 AVMSJ5	Ig heavy chain v r
34	5	25.0	122	2 T14821	aldehyde dehydroge
35	5	25.0	124	2 G82745	hypothetical prote
36	5	25.0	130	2 T11322	NADH2 dehydrogenas
37	5	25.0	130	2 S06817	Ig heavy chain v r
38	5	25.0	134	2 E70251	hypothetical prote
39	5	25.0	136	2 S13791	Ig heavy chain v r
40	5	25.0	137	2 T12880	hypothetical prote
41	5	25.0	139	1 RKOMSY	ribulose-bisphosph
42	5	25.0	139	2 A25912	Ig heavy chain pre
43	5	25.0	155	2 A13130	hypothetical prote
44	5	25.0	163	1 B39415	fimbrial protein s
45	5	25.0	166	2 JCI295	papF fimbrial prot

ALIGNMENTS

RESULT 1

GNWVC3

genome polyprotein - hepatitis C virus (strain HCV-1)

N;Contains: capsid protein C; envelope protein M; hepatitis M; hepatitis C virus (strain HCV-1) protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C;Species: hepatitis C virus

C;Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 19-Jan-2001

C;Accession: A39166; PQ0403; PQ0404

R;Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Co

Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991

A;Title: Genetic organization and diversity of the hepatitis C virus.

A;Reference number: A39166; MUID:91172826; PMID:1848704

A;Accession: A39166

A;Molecule type: mRNA

A;Residues: 1-3011 <CHA>

A;Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874

R;Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.

J. Gen. Virol. 73, 1131-1141, 1992

A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to

A;Reference number: PQ0393; MUID:92268871; PMID:1316939

A;Accession: PQ0403

A;Molecule type: Genomic RNA

A;Residues: 1577-1633 <CHA>

A;Cross-references: DDBJ:D10128

A;Experimental source: isolates E-b16

A;Accession: PQ0404

A;Status: preliminary

A;Molecule type: Genomic RNA

A;Residues: 1577-1633 <CH2>

A;Experimental source: isolates E-b17

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructur

F;1-115/Product: capsid protein C #status: predicted <CPC>

F;116-191/Product: envelope protein M #status: predicted <EPM>

F;192-389/Product: major envelope protein E #status: predicted <MEE>

F;390-729/Product: nonstructural protein NS1 #status: predicted <NS1>

F;730-1006/Product: nonstructural protein NS2 #status: predicted <NS2>

F;1007-1615/Product: hepatitis C virus genome polyprotein

F;1230-1337/Region: nucleotide-binding motif A (P-loop)

F;1312-1317/Region: nucleotide-binding motif B

F;1316-1319/Region: DEXH motif

F;1616-1862/Product: nonstructural protein NS4a #status: predicted <N4A>

F;1863-2013/Product: nonstructural protein NS5 #status: predicted <N4B>

F;2014-3011/Product: nonstructural protein NS5 #status: predicted <NS5>

F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,2

Query Match 100.0%; Score 20; DB 1; Length 3011;

Best Local Similarity 100.0%; Pred. No. 3.2e-15;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPRYLWHYPCTINYTI 20

|||||

Db 607 CLVDYPYRLWHYPCTINYTI 626

RESULT 2

S40770 genome polyprotein - hepatitis C virus
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C;Accession: S40770; P1285
R;Okamoto, H.
submitted to the EMBL Data Library, March 1992
A;Reference number: S40770
A;Accession: S40770
A;Molecule type: genomic RNA
A;Residues: 1-3011 <OK>
A;Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990
A;Title: The 5'-terminal sequence of the hepatitis C virus genome.
A;Reference number: P1284; MUID:91013116; PMID:2170712
A;Accession: P1285
A;Molecule type: genomic RNA
A;Residues: 1-513 <OK2>
A;Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
A;Experimental source: isolate HC-J1
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPW>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;730-1006/Product: nonstructural protein NS1 #status predicted <NS1>
F;1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>
F;1230-1237/Region: hepatitis C virus genome
F;1312-1317/Region: nucleotide-binding motif A (P-loop)
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.2e-15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20

607 CLVDYPYRLWHYPCTINYTI 626

RESULT 3

GNMVCH genome polyprotein - hepatitis C virus (strain H)
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Accession: A36814; A41546
R;Inchauspe, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A;Description: Genomic structure of the human prototype strain H of hepatitis C virus: c
A;Reference number: A36814
A;Accession: A36814
A;Molecule type: genomic RNA
A;Residues: 1-3011 <INC>
A;Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
R;Inchauspe, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari
A;Reference number: A41546; MUID:92052256; PMID:1658800
A;Contents: annotation

A;Note: neither amino acid nor nucleotide sequence is given
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructur
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPW>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;730-1006/Product: nonstructural protein NS1 #status predicted <NS1>
F;1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>
F;1230-1237/Region: hepatitis C virus genome
F;1312-1317/Region: nucleotide-binding motif A (P-loop)
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F;196.209,234,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,2

Query Match 90.0%; Score 18; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 6.9e-13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VDPYRLWHYPCTINYTI 20

609 VDPYRLWHYPCTINYTI 626

RESULT 4

S35631 genome polyprotein - hepatitis C virus (fragment)
N;Contains: nonstructural protein 1
C;Species: hepatitis C virus
C;Date: 19-Mar-1997 #sequence_revision 30-Jan-1998 #text_change 17-Nov-2000
C;Accession: S35631
R;Sarashina, T.; Sakurai, T.; Watanabe, Y.; Kashima, K.; Suzuki, T.; Chiba, Y.; Kita, Y.
Nucleic Acids Res. 21, 1037, 1993
A;Title: Nucleotide sequence of the hepatitis C virus genome from a patient negative fo
A;Reference number: S35629; MUID:93197128; PMID:8383835
A;Accession: S35631
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-350 <SAR>
A;Cross-references: EMBL:D13970
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1992
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: polyprotein
F;1-350/Product: nonstructural protein 1 #status predicted <MAT>

Query Match 75.0%; Score 15; DB 2; Length 350;
Best Local Similarity 100.0%; Pred. No. 3.5e-10;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCT 15

228 CLVDYPYRLWHYPCT 242

RESULT 5

JQ1584 genome polyprotein - hepatitis C virus (strain U.K.) (fragment)
N;Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural pro
C;Species: hepatitis C virus
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
C;Accession: JQ1584
R;Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.
J. Gen. Virol. 73, 1521-1525, 1992
A;Title: Cloning and sequencing of the structural region and expression of putative cor
A;Reference number: JQ1584; MUID:92300349; PMID:1318944
A;Accession: JQ1584
A;Molecule type: genomic RNA
A;Residues: 1-640 <KUM>
A;Cross-references: GB:X84079; NID:g643119; PIDN:CAA58888.1; PID:g643120
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; polypr

F:1-191/Product: core protein C #status predicted <CPC>
 F:192-389/Product: envelope protein E1 #status predicted <BE1>
 F:390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predicted <NS1>
 F:196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn) (cova

Query Match 75.0%; Score 15; DB 2; Length 640;
 Best Local Similarity 100.0%; Pred. No. 5.9e-10;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCT 15
 |||||
 Db 607 CLVDYPYRLWHYPCT 621

RESULT 6
 GNMVTC
 genome polyprotein - hepatitis C virus (strain HC-J8)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 F:1-191/Product: core protein C #status predicted <MAT1>
 F:192-389/Product: envelope protein E1 #status predicted <MAT2>
 F:390-640/Product: envelope protein E2 #status predicted <MAT3>
 F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

C:Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text_change 19-Jan-2001
 C:Accession: A40250; PQ0397; PQ0559
 R:Okamoto, H.; Kura, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.;
 Virol. 189, 331-341, 1992
 A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to rep
 A:Reference number: A40250; MUID:92230232; PMID:1314459
 A:Accession: A40250
 A:Molecule type: Genomic RNA
 A:Residues: 1-3033 <OK>
 A:Cross-references: GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1; PID:g221609
 J.Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.I
 J. Gen. Virol. 73, 1131-1141, 1992
 A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
 A:Reference number: PQ0393; MUID:92268871; PMID:1316939
 A:Accession: PQ0397
 A:Molecule type: Genomic RNA
 A:Residues: 2678-2754 <CHA>
 A:Cross-references: DBJ:D10134
 A:Experimental source: isolate E-b12
 R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno
 Biochem. Biophys. Res. Commun. 191, 279-285, 1991
 A:Title: Distribution of plural HCV types in Japan.
 A:Reference number: PQ0559; MUID:92068204; PMID:1720309
 A:Accession: PQ0559
 A:Molecule type: mRNA
 A:Residues: 2678-2729 <KAT>
 A:Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <ME>
 F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
 F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1011-1619/Product: hepatitis C virus genome polyprotein
 F:1234-1241/Region: nucleotide-binding motif A (P-loop)
 F:1316-1321/Region: nucleotide-binding motif B
 F:1320-1323/Region: DEXH motif
 F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23

Query Match 75.0%; Score 15; DB 1; Length 3033;
 Best Local Similarity 100.0%; Pred. No. 2.2e-09;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCT 15
 |||||
 Db 611 CLVDYPYRLWHYPCT 625

RESULT 7

S19875
 genome polyprotein - hepatitis C virus (isolate JK3) (fragment)
 N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
 C:Species: hepatitis C virus
 A:Variety: isolate JK3
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
 C:Accession: S19875
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus
 A:Reference number: S18029
 A:Accession: S19875
 A:Molecule type: Genomic RNA
 A:Residues: 1-782 <HON>
 A:Cross-references: EMBL:X61592; NID:g59482; PIDN:CAA43789.1; PID:g59483
 A:Experimental source: isolate JK3
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
 F:1-191/Product: core protein #status predicted <MAT1>
 F:192-383/Product: envelope protein 1 #status predicted <MAT2>
 F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
 F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 65.0%; Score 13; DB 2; Length 782;
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VDYPYRLWHYPCT 15
 |||||
 Db 609 VDYPYRLWHYPCT 621

RESULT 8
 PNO677
 hypothetical protein 787 - hepatitis C virus (fragment)
 C:Species: hepatitis C virus
 C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000
 C:Accession: PNO677
 R:Cho, S.H.; Yoon, J.I.; Chang, J.E.; Ahn, B.M.; Lee, C.H.; Lee, Y.I.
 Biochem. Biophys. Res. Commun. 196, 780-788, 1993
 A:Title: Genomic typing of hepatitis C viruses from Korean patients: Implications of ge
 A:Reference number: PNO677; MUID:94059104; PMID:8240354
 A:Accession: PNO677
 A:Molecule type: mRNA
 A:Residues: 1-787 <CHO>
 A:Cross-references: GB:L20498; NID:g1381031; PIDN:AAB02608.1; PID:g1381032
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: glycoprotein; nonstructural protein
 F:196,209,234,250,305,325,421,427,452,536,544,560,580,627,649/Binding site: carbohydrate

Query Match 65.0%; Score 13; DB 2; Length 787;
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VDYPYRLWHYPCT 15
 |||||
 Db 613 VDYPYRLWHYPCT 625

RESULT 9
 GNMVTC
 genome polyprotein - hepatitis C virus
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
 C:Accession: A38465
 R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.;
 J. Virol. 65, 1105-1113, 1991
 A:Title: Structure and organization of the hepatitis C virus genome isolated from human
 A:Reference number: A38465; MUID:91140698; PMID:1847440
 A:Accession: A38465
 A:Molecule type: genomic RNA

A>Note: this sequence is inconsistent with the nucleotide translation
A>Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320 as Trp, and TTC for residue 771 as Ser
A>Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBIP:121748)

C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine
F:116-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (Asn)

Query Match 65.0%; Score 13; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 4.9e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VDYPRLWHYPCT 15

Db 609 VDYPRLWHYPCT 621

RESULT 13

GNWVTW
genome polyprotein - hepatitis C virus (strain Taiwan)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40244
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992
A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
A:Reference number: A40244; MUID:92230206; PMID:1314449

A:Accession: A40244
A:Molecule type: genomic RNA
A:Residues: 1-3010 <CHE>

A:Cross-references: GB:M84754
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F:116-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,417,423,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 65.0%; Score 13; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 4.9e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VDYPRLWHYPCT 15

Db 609 VDYPRLWHYPCT 621

RESULT 14

JQ1303

genome polyprotein - hepatitis C virus (isolate HC-J6)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
C:Accession: JQ1303
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.
J. Gen. Virol. 72, 2697-2704, 1991
A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a hu
A:Reference number: JQ1303; MUID:92044440; PMID:1658196

A:Accession: JQ1303

A:Molecule type: genomic RNA
A:Residues: 1-3033 <OKA>

A:Cross-references: GB:D00944; NID:G221650; PIDN:BAA00792.1; PID:G221651
A:Experimental source: isolate HC-J6 from a Japanese individual
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; trans
F:116-191/Product: capsid protein C #status predicted <CPC>
F:192-389/Product: major envelope protein M #status predicted <EPM>
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F:1011-1619/Product: hepatitis C virus genome polyprotein
F:1316-1321/Region: nucleotide-binding motif B
F:1320-1323/Region: DEXH motif
F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,2

Query Match 60.0%; Score 12; DB 1; Length 3033;
Best Local Similarity 100.0%; Pred. No. 7.3e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 DYPRLWHYPCT 15

Db 614 DYPRLWHYPCT 625

RESULT 15

JQ1366
polyprotein - hepatitis C virus (French isolate) (fragments)
C:Species: hepatitis C virus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C:Accession: JQ1366
R:Krensdorf, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.
J. Gen. Virol. 72, 2557-2561, 1991
A:Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication
A:Reference number: JQ1366; MUID:92013977; PMID:1655961

A:Accession: JQ1366

A:Molecule type: genomic RNA
A:Residues: 1-716 <KRE>

C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: glycoprotein; polyprotein

F:84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #sta
Query Match 55.0%; Score 11; DB 2; Length 716;
Best Local Similarity 100.0%; Pred. No. 3.1e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 YPRLWHYPCT 15

Db 278 YPRLWHYPCT 288

Search completed: November 21, 2003, 21:11:37
Job time : 9.25 secs

THIS PAGE BLANK (11SP70)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:52 ; Search time 4.85 Seconds
(without alignments)
193.925 Million cell updates/sec

Title: US-09-973-025-88

Perfect score: 20

Sequence: 1 CLVDYPLRLWHPCTINYTI 20

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	3011	1 POLG_HCV1	P26664 h genome po
2	18	90.0	3011	1 POLG_HCVH	P27958 h genome po
3	15	75.0	737	1 POLG_HCVJ7	P27961 hepatitis c
4	15	75.0	3033	1 POLG_HCVJ8	P26661 h genome po
5	13	65.0	737	1 POLG_HCVJ5	P27960 hepatitis c
6	13	65.0	3010	1 POLG_HCVBK	P26663 h genome po
7	13	65.0	3010	1 POLG_HCVJA	P26662 h genome po
8	13	65.0	3010	1 POLG_HCVJT	Q00269 h genome po
9	13	65.0	3010	1 POLG_HCVTW	P29846 h genome po
10	12	60.0	3033	1 POLG_HCVJ6	P26660 h genome po
11	6	30.0	265	1 YOR3_LELV	Q04567 lelystad vi
12	6	30.0	312	1 TRUB_HAEIN	P45142 haemophilus
13	5	25.0	89	1 CYB_BRANA	P49390 brassica na
14	5	25.0	116	1 HV36_MOUSE	P01806 mus musculu
15	5	25.0	117	1 HV41_MOUSE	P01811 mus musculu
16	5	25.0	117	1 HV42_MOUSE	P01812 mus musculu
17	5	25.0	118	1 HV39_MOUSE	P01809 mus musculu
18	5	25.0	119	1 HV37_MOUSE	P01807 mus musculu
19	5	25.0	119	1 HV38_MOUSE	P01808 mus musculu
20	5	25.0	119	1 HV40_MOUSE	P01810 mus musculu
21	5	25.0	139	1 RBS_CYLSN	P24683 cylindrothe
22	5	25.0	167	1 PAPF_ECOLI	P08408 escherichia
23	5	25.0	167	1 PRGF_ECOLI	P42187 escherichia
24	5	25.0	169	1 LOJA_CAMJE	Q9paz0 campylobact
25	5	25.0	208	1 YQJ9_CABEL	Q95216 caenorhabdi
26	5	25.0	209	1 HAM1_CHLMU	Q997d4 chlamydia m
27	5	25.0	209	1 HAM1_CHLTR	O84611 chlamydia t
28	5	25.0	211	1 HPRT_LEIDO	P43152 leishmania
29	5	25.0	216	1 CSGD_ECOLI	P52106 escherichia
30	5	25.0	216	1 CSGD_SALTY	O54294 salmonella
31	5	25.0	228	1 BIOD_CLOPE	Q8XK60 clostridium
32	5	25.0	228	1 D7A1_RAT	Q64057 rattus norv
33	5	25.0	236	1 VC05_SPVKA	P32227 swinepox vi

RESULT 1

ID	POLG_HCV1	STANDARD;	PRT;	3011 AA.
AC	P26664;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].			
OS	Hepatitis C virus (isolate 1) (HCV).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.			
OX	NCBI_TaxID=11104;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91172826; PubMed=1848704;			
RA	Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;			
RA	"Genetic organization and diversity of the hepatitis C virus."; Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).			
RL	-!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.			
CC	NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.			
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	-!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA}(N).			
CC	-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.			
CC	-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; M62321; AAA45676.1; .			
DR	PIR; A39166; GNWVC3.			
DR	PDB; 1A1V; 16-FEB-99.			
DR	PDB; 1HEI; 25-NOV-98.			
DR	MEROPS; S29.001; .			
DR	MEROPS; U39.001; .			
DR	InterPro; IPR001410; DEAD.			
DR	InterPro; IPR002522; HCV_capsid.			

Q92170 helicobacte
P56009 helicobacte
Q9pad8 xytella fas
P72988 synechocyst
P38506 escherichia
O32153 bacillus su
O75123 homo sapien
Q9piz5 campylobact
P23820 todarodes p
Q58290 methanococc
Q58686 methanococc
P43356 homo sapien

ALIGNMENTS

DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR007095; RNA_pol_Ds_Ps.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; Helicase C; 1.
 DR Pfam; PF00998; Viral_RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3011
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 476 476
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2364 2364
 FT CARBOHYD 2789 2789
 FT CARBOHYD 3011 3011
 FT SEQUENCE 3011 AA; 327197 MW; 65F8C9447FC5AF9 CRC64;
 Query Match 100.0%; Score 20; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 1.2e-14;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 607 CLVDYPYRLWHYPTINYTI 626
 RESULT 2
 ID POLG_HCVH STANDARD; PRT; 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 RN NCBI_TaxID=11108;
 RX [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=1658800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 RL Structure 6:89-100(1998).
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M67463; AAA45534.1; -
DR PIR; A36814; GNWVCH.
DR PDB; 1HEI; 25-NOV-98.
DR PDB; 1AIV; 16-FEB-99.
DR PDB; 1AIR; 17-JUN-98.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR TRANSFAC; T04155; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA pol DS ps.
DR InterPro; IPR007094; RNA pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
FT CHAIN 192 383 CAPSID PROTEIN C.
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
FT CHAIN 810 1026 PROTEIN P7.
FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5A.
FT CHAIN 3011 369 NONSTRUCTURAL PROTEIN NS5B.
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT PTE BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT STRAND 1224 1226
FT TURN 1232 1233
FT TURN 1236 1238
FT HELIX 1239 1246
FT TURN 1247 1248
FT STRAND 1251 1255
FT HELIX 1258 1271
FT TURN 1272 1272
FT STRAND 1277 1280
FT TURN 1281 1282
FT STRAND 1283 1285
FT STRAND 1291 1295
FT HELIX 1296 1301
FT TURN 1302 1303
FT STRAND 1312 1316
FT TURN 1317 1319
FT HELIX 1323 1335
FT TURN 1336 1340
FT STRAND 1343 1347
FT TURN 1352 1353
FT TURN 1360 1361
FT STRAND 1362 1366
FT STRAND 1368 1368
FT STRAND 1372 1375
FT TURN 1376 1377
FT STRAND 1378 1380
FT HELIX 1382 1385
FT STRAND 1389 1393
FT HELIX 1397 1409
FT TURN 1410 1411
FT STRAND 1414 1417
FT TURN 1419 1420
FT STRAND 1432 1436
FT TURN 1438 1439
FT STRAND 1450 1453
FT STRAND 1456 1463
FT STRAND 1471 1478
FT STRAND 1480 1480
FT HELIX 1481 1488
FT TURN 1489 1490
FT STRAND 1497 1501
FT STRAND 1507 1507
FT STRAND 1511 1511
FT HELIX 1514 1527
FT HELIX 1532 1544
FT STRAND 1550 1550
FT HELIX 1555 1564
FT HELIX 1570 1578
FT TURN 1579 1580
FT TURN 1584 1597
FT TURN 1598 1598
FT HELIX 1606 1611
FT TURN 1614 1618
FT STRAND 1622 1623
FT STRAND 1627 1627
FT STRAND 1635 1636
FT HELIX 1640 1652
SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;
Query Match 90.0%; Score 18; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.1e-12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 VDYPYRLWHYPTINVTI 20
Db 609 VDYPYRLWHYPTINVTI 626
RESULT 3
POLG_HCVJ7
ID POLG_HCVJ7 STANDARD; PRT; 737 AA.
AC P27961;
DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
 DE protein (Envelope protein M); Major envelope protein E; Nonstructural
 DE proteins NS1 and NS2] (Fragment).
 OS Hepatitis C virus (isolate HC-J7) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11114;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D10077; BAA00971.1; -;
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV env; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR PolyProtein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT INIT_MET 1 1
 FT REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT NON_TER 737 737
 SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122P2 CRC64;

Query Match 75.0%; Score 15; DB 1; Length 737;
 Best Local Similarity 100.0%; Pred. No. 1.4e-09; Mismatches 0; Gaps 0;
 Matches 15; Conservative 0; Indels 0;
 QY 1 CLVDYPYRLWHYPCT 15
 DB 611 CLVDYPYRLWHYPCT 625
 RESULT 4
 POLG HCVJ8
 ID POLG HCVJ8 STANDARD; PRT; 3033 AA.
 AC P26661;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP58) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J8) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11115;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
 RA Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA) (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D10988; BAA01761.1; -;
 DR PIR; A40250; GNWJ8.
 DR HSSP; P27958; 1HEI.
 DR MEROPS; S29.001; -;
 DR MEROPS; U39.001; -;
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.


```

FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1031 1035 HELIX 1047
FT STRAND 1050 1050 STRAND
FT STRAND 1059 1063 STRAND
FT STRAND 1068 1074 STRAND
FT TURN 1075 1076 HELIX 1077
FT STRAND 1082 1085 STRAND
FT HELIX 1086 1087 HELIX 1088
FT TURN 1089 1092 STRAND
FT TURN 1093 1094 STRAND
FT STRAND 1095 1097 STRAND
FT STRAND 1101 1103 STRAND
FT TURN 1104 1107 STRAND
FT STRAND 1108 1112 STRAND
FT STRAND 1120 1120 STRAND
FT STRAND 1122 1122 STRAND
FT STRAND 1129 1133 STRAND
FT TURN 1135 1136 STRAND
FT STRAND 1139 1144 STRAND
FT STRAND 1149 1157 STRAND
FT HELIX 1158 1161 STRAND
FT TURN 1162 1163 STRAND
FT TURN 1165 1166 STRAND
FT STRAND 1168 1171 STRAND
FT TURN 1172 1174 STRAND
FT STRAND 1175 1186 STRAND
FT TURN 1187 1188 STRAND
FT STRAND 1189 1197 STRAND
FT HELIX 1198 1202 STRAND
FT TURN 1203 1204 STRAND
FT STRAND 1680 1688 STRAND
SQ SEQUENCE 3010 AA; 327189 MW; F8422D58CCFFDF9C CRC64;

Query Match 65.0%; Score 13; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 7.6e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VDYPRLWHYPCT 15
DB 609 VDYPRLWHYPCT 621

RESULT 7
POLG_HCVVA STANDARD; PRT; 3010 AA.
AC P26662;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11116;
RN [1]
RP SEQUENCE FROM N.A.

```

MEDLINE=91089550; PubMed=2175903;
 Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 Sugimura T., Shimotohno K.;
 "Molecular cloning of the human hepatitis C virus genome from
 Japanese patients with non-A, non-B hepatitis";
 Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 [2]
 DISCUSSION OF SEQUENCE.
 MEDLINE=91192160; PubMed=1849488;
 Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., MuraIso K.,
 Ohkoshi S., Shimotohno K.;
 "Molecular structure of the Japanese hepatitis C viral genome.";
 FEBS Lett. 280:325-328(1991).
 -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 {RNA} (N).
 -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA.
 -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; D90208; BAAL4233.1; -;
 PIR; A39253; GNWVCJ.
 HSP; P26663; LJXP.
 MEROPS; S29.001; -;
 MEROPS; U39.001; -;
 InterPro; IPR001410; DEAD.
 InterPro; IPR002522; HCV_capsid.
 InterPro; IPR002521; HCV core.
 InterPro; IPR002519; HCV env.
 InterPro; IPR002531; HCV NS1.
 InterPro; IPR002518; HCV NS2.
 InterPro; IPR004109; HCV NS3.
 InterPro; IPR000745; HCV NS4a.
 InterPro; IPR001490; HCV NS4b.
 InterPro; IPR002868; HCV NS5a.
 InterPro; IPR002166; HCV RdRP.
 InterPro; IPR001650; Helicase C.
 InterPro; IPR007095; RNA_pol_PS.
 InterPro; IPR007094; RNA_pol_Psvir.
 Pfam; PF01543; HCV_capsid; 1.
 Pfam; PF01542; HCV_core; 1.
 Pfam; PF01539; HCV_env; 1.
 Pfam; PF01560; HCV_NS1; 1.
 Pfam; PF01538; HCV_NS2; 1.
 Pfam; PF02907; HCV_NS3; 1.
 Pfam; PF01006; HCV_NS4a; 1.
 Pfam; PF01001; HCV_NS4b; 1.
 Pfam; PF01506; HCV_NS5a; 1.
 Pfam; PF00271; Helicase C; 1.
 Pfam; PF00998; Viral_RdRP; 1.
 ProDom; PD186062; HCV_NS1; 1.
 SMART; SM00487; DEXDC; 1.
 PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 INIT_MET 1 1
 CHAIN 1 115

FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).

FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).

FT CHAIN 384 729 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).

FT CHAIN 730 1006 NON-STRUCTURAL PROTEIN NS1 (POTENTIAL).

FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).

FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).

FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).

FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).

FT TRANSMEM 347 369

FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT NP_BIND 1230 1237 ATP (POTENTIAL).

FT SITE 1316 1319 DECH BOX.

FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;

Query Match 65.0%; Score 13; DB 1; Length 3010;

Best Local Similarity 100.0%; Pred. No. 7.6e-07;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 VDYPYRLWHYPCT 15
|||||

Db 609 VDYPYRLWHYPCT 621

RESULT 8

ID POLG HCVJT STANDARD; PRT; 3010 AA.

AC Q00269;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP35); Envelope glycoprotein E2 (GP68) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P36); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

OS Hepatitis C virus (isolate HC-JT) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus

OX NCBI_TaxID=31642;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92295714; PubMed=1318627;

RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J., Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;

RA "Molecular cloning of hepatitis C virus genome from a single Japanese carrier: sequence variation within the same individual and among infected individuals."

RT Virus Res. 23:39-53(1992).

CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

CC EMBL; D11168; BA01943.1; -.

DR PIR; A45573; A45573.

DR PDB; 1AIQ; 25-MAR-98.

DR PDB; 1JXP; 14-JAN-98.

DR MEROPS; S29, 001; -.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV capsid.

DR InterPro; IPR002521; HCV core.

DR InterPro; IPR002519; HCV env.

DR InterPro; IPR002531; HCV NS1.

DR InterPro; IPR002518; HCV NS2.

DR InterPro; IPR004109; HCV NS3.

DR InterPro; IPR000745; HCV NS4a.

DR InterPro; IPR001490; HCV NS4b.

DR InterPro; IPR002868; HCV NS5a.

DR InterPro; IPR002166; HCV RdRP.

DR InterPro; IPR007095; RNA_pol_DS_PS.

DR InterPro; IPR007094; RNA_pol_PSVir.

DR Pfam; PF01543; HCV_capsid; 1.

DR Pfam; PF01542; HCV_core; 1.

DR Pfam; PF01539; HCV_env; 1.

DR Pfam; PF01560; HCV_NS1; 1.

DR Pfam; PF01538; HCV_NS2; 1.

DR Pfam; PF02307; HCV_NS3; 1.

DR Pfam; PF01006; HCV_NS4a; 1.

DR Pfam; PF01001; HCV_NS4b; 1.

DR Pfam; PF01506; HCV_NS5a; 1.

DR Pfam; PF00271; helicase_C; 1.

DR Pfam; PF00998; Viral_RdRP; 1.

DR ProDom; PD186062; HCV_NS1; 1.

DR SMART; SM00487; DEXDC; 1.

DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase; Core protein; Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane; Nonstructural protein; Hydrolase; Serine protease; 3D-structure.

FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE.

FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).

FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).

FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).

FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).

FT CHAIN 730 1006 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).

FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).

FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).

FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).

FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).

FT TRANSMEM 347 369 POTENTIAL.

FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT NP_BIND 1230 1237 ATP (POTENTIAL).

FT SITE 1316 1319 DECH BOX.

FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;
 Query Match 65.0%; Score 13; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 7.6e-07;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 VDPYRLWHPCT 15
 DB 609 VDPYRLWHPCT 621
 RESULT 9
 POLG_HCVTW STANDARD; PRT: 3010 AA.
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230206; PubMed=1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 RT "The Taiwanese hepatitis C virus genome: sequence determination and
 RT mapping the 5' termini of viral genomic and antigenomic RNA.";
 RL Virology 188:102-113(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; M84754; -; NOT_ANNOTATED_CDS.
 PIR; A40244; GNVVTV.
 PDB; 1M64; 25-FEB-03.
 PDB; 1NS3; 08-APR-98.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002152; HCV capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV NS5b.
 DR InterPro; IPR007095; RNA pol DS PS.
 DR InterPro; IPR007094; RNA pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; helicase C; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CORE PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 3010 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

```
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 65.0%; Score 13; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 7.6e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VDPYRLWHYPCT 15
DB 609 VDPYRLWHYPCT 621

RESULT 10
POLG_HCVJ6 STANDARD; PRT; 3033 AA.
AC P26660;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P66); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P67) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J6) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11113;
RX MEDLINE=9204440; PubMed=1659196;
RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
RA Machida A., Miyakawa Y., Mayumi M.;
RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704 (1991).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; D00944; BAA00792.1; --
DR PIR; JQ1303; JQ1303.
DR HSP; P27958; IHEI.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
```

```
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV_Ns1.
DR InterPro; IPR002518; HCV_Ns2.
DR InterPro; IPR004109; HCV_Ns3.
DR InterPro; IPR00745; HCV_Ns4a.
DR InterPro; IPR001490; HCV_Ns4b.
DR InterPro; IPR002868; HCV_Ns5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA pol DS ps.
DR InterPro; IPR007094; RNA pol psvir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV_Ns1; 1.
DR Pfam; PF01538; HCV_Ns2; 1.
DR Pfam; PF02907; HCV_Ns3; 1.
DR Pfam; PF01006; HCV_Ns4a; 1.
DR Pfam; PF01001; HCV_Ns4b; 1.
DR Pfam; PF01506; HCV_Ns5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral RdRp; 1.
DR ProDom; PD186062; HCV_Ns1; 1.
DR SMART; SM00487; DEXDC; 1.
DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolyase; Serine protease.
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 131
FT CHAIN 132 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT TRANSMEM 347 369
FT ACT_SITE 1087 1087
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT NP_BIND 1234 1241
FT SITE 1320 1323
FT CARBOHYD 136 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT CARBOHYD 1091 1091
FT CARBOHYD 2038 2038
FT CARBOHYD 2811 2811
SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;

Query Match 60.0%; Score 12; DB 1; Length 3033;
Best Local Similarity 100.0%; Pred. No. 9.9e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 DYPYRLWHYPCT 15
DB 614 DYPYRLWHYPCT 625
```

```

RESULT 11
YOR3_DELV
ID YOR3_DELV STANDARD; PRT; 265 AA.
AC Q04567;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Hypothetical 30.6 kDa protein (ORF3).
OS Lelystad virus (LV) (porcine reproductive and respiratory syndrome virus).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID=11049;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93297139; PubMed=8517032;
RA Meulenbergh J.J.M., Hulst M.M., de Meijer E.J., Moonen P.L.J.M.,
RA den Besten A., de Kluiver E.P., Wensvoort G., Moormann R.J.M.;
RT "Lelystad virus, the causative agent of porcine epidemic abortion and
RT respiratory syndrome (PEARS), is related to LDV and EAV.";
RL Virology 192:62-72(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Boxmeer 10;
RX MEDLINE=93174942; PubMed=8438574;
RA Conzelmann K.K., Visser N., van Woensel P., Thiel H.J.;
RT "Molecular characterization of porcine reproductive and respiratory
RT syndrome virus, a member of the arterivirus group.";
RL Virology 193:329-339(1993).
CC -!- FUNCTION: POSSIBLE ENVELOPE PROTEIN.
CC -!- SIMILARITY: TO THE LDV CORRESPONDING PROTEIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M96262; AAA46276.1; -
DR EMBL; L04493; AAA47103.1; -
DR PIR; C45392; C45392.
DR PIR; D36861; D36861.
DR InterPro; IPR002556; Arteri env.
DR Pfam; PF01606; Arteri env; 1.
DR ProDom; PD002543; Arteri env; 1.
KW Hypothetical protein, Envelope protein.
FT CONFLICT 3 5 HQC -> ROR (IN REF. 2).
FT CONFLICT 150 150 H -> P (IN REF. 2).
SQ SEQUENCE 265 AA; 30583 MW; B35A13D7EBB5D51C CRC64;

Query Match 30.0%; Score 6; DB 1; Length 265;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 TINYTI 20
DB 48 TINYTI 53

RESULT 12
TRUB_HAEIN
ID TRUB_HAEIN STANDARD; PRT; 312 AA.
AC P45142;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE tRNA pseudouridine synthase B (EC 4.2.1.70) (tRNA pseudouridine 55
DE synthase) (Psi55 synthase) (Pseudouridylylase)
DE hydrolyase).
GN TRUB OR H11289.
OS Haemophilus influenzae.

```

```

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: FORMATION OF PSEUDOURIDINE AT POSITION 55 IN THE
CC PSI GC LOOP OF TRANSFER RNAs (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine
CC 5'-phosphate + H(2)O.
CC -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U32809; AAC22938.1; -
DR PIR; B64170; B64170.
DR TIGR; H11289; -
DR InterPro; IPR004510; TruB.
DR InterPro; IPR002501; TruB_N.
DR Pfam; PF01509; TruB_N; 1.
DR TIGRPFAM; TIGR00431; TruB; 1.
KW Lyase; tRNA processing; Complete proteome.
SQ SEQUENCE 312 AA; 35085 MW; C423FAD92B6E0295 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 TINYTI 20
DB 299 TINYTI 304

RESULT 13
CYB_BRANA
ID CYB_BRANA STANDARD; PRT; 89 AA.
AC P49390;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cytochrome b (fragment).
GN MTCYB OR COB OR CYTB.
OS Brassica napus (Rape).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RX MEDLINE=94073988; PubMed=8252643;
RA Ye F., Bernhardt J., Abel W.O.;
RT "Genes for ribosomal proteins S3, L16, L5 and S14 are clustered in

```

RT the mitochondrial genome of *Brassica napus* L.;"

RL Curr. Genet. 24:323-329(1993).

CC -!- FUNCTION: Component of the ubiquinol-cytochrome c reductase

CC complex (complex III or cytochrome b-c1 complex), which is a

CC respiratory chain that generates an electrochemical potential

CC coupled to ATP synthesis (By similarity).

CC -!- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or

CC b562) is low-potential and absorbs at about 562, and heme 2 (or BH

CC or b566) is high-potential and absorbs at about 566 (By

CC similarity).

CC -!- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,

CC cytochrome c1 and the Rieske protein (By similarity).

CC -!- SIMILARITY: Belongs to the cytochrome b family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: X68727; CAA48669.1; ..

CC InterPro: IPR005798; Cytb_b6_C.

CC InterPro: IPR005797; Cytb_b6_N.

CC Pfam: PF00033; cytochrome b_N; 1.

CC PROSITE: PS00192; CYTOCHROME B HEME; 1.

CC PROSITE: PS00193; CYTOCHROME B_OO; PARTIAL.

CC Electron transport; Mitochondrion; Respiratory chain; Transmembrane;

CC Heme.

CC METAL 88 88 IRON 1 (HEME B562 AXIAL LIGAND).

CC NON_TER 89 89

CC SEQUENCE 89 AA; 10215 MW; 9924A3BABF6DAB61 CRC64;

Query Match 25.0%; Score 5; DB 1; Length 89;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVDYP 6

DB 23 LVDYP 27

RESULT 14

HV36 MOUSE

ID HV36 MOUSE STANDARD; PRT; 116 AA.

AC P01806;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DE Ig heavy chain V region 441 precursor.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=82059449; PubMed=6795591;

RA Ollio R., Auffray C., Sikorav J.-L., Rougeon F.;

RT "Mouse heavy chain variable regions: nucleotide sequence of a

RT germ-line Vh gene segment.";

RL Nucleic Acids Res. 9:4099-4109(1981).

CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: V00774; CAA24152.1; ..

DR PIR; A02076; HVMS44.

DR HSSP; P01810; 2FBU.

DR MGD; MG1:96492; Igh-VX24.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003006; Ig MHC.

DR InterPro; IPR003596; Ig_v.

DR Pfam; PF00047; ig; 1.

DR SMART; SM00406; IGV; 1.

DR PROSITE; PS50835; IG-LIKE; 1.

KW Immunoglobulin V region; Signal.

FT CHAIN 1 18 IG HEAVY CHAIN V REGION 441.

FT CHAIN 19 116 IG-LIKE.

FT DOMAIN 19 >116

FT NON_TER 116

SQ SEQUENCE 116 AA; 12911 MW; 6B4FF3EF9A95D9F3 CRC64;

Query Match 25.0%; Score 5; DB 1; Length 116;

Best Local Similarity 100.0%; Pred. No. 41;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 TINYT 19

DB 75 TINYT 79

RESULT 15

HV41 MOUSE

ID HV41 MOUSE STANDARD; PRT; 117 AA.

AC P01811;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DE Ig heavy chain V region UPC10.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=83021113; PubMed=6181731;

RA Auffray C., Sikorav J.-L., Ollio R., Rougeon F.;

RT "Correlation between D region structure and antigen-binding

RT specificity: evidences from the comparison of closely related

RT immunoglobulin VH sequences.";

RL Ann. Immunol. (Paris) 132D:77-88(1981).

CC -!- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IG GAMMA-2A MYELOMA

CC PROTEIN BINDING 2,6-LEVAN.

CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: M36631; AAA38078.1; ..

DR PIR; A02081; G2MSU1.

DR HSSP; P01810; 2FBU.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003006; Ig MHC.

DR InterPro; IPR003596; Ig_v.

DR Pfam; PF00047; ig; 1.

DR SMART; SM00406; IGV; 1.

DR PROSITE; PS50835; IG-LIKE; 1.

KW Immunoglobulin V region.

FT DOMAIN 1 116 IG-LIKE.

FT NON_TER 117

SQ SEQUENCE 117 AA; 13001 MW; B20A1074F8E99E7F CRC64;

Query Match 25.0%; Score 5; DB 1; Length 117;

Best Local Similarity 100.0%; Pred. No. 41;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 TINYT 19
|||||
Db 57 TINYT 61

Search completed: November 21, 2003, 21:00:02
Job time : 5.85 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:45:57 ; Search time 24.4 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-88
Perfect score: 20
Sequence: 1 CLVDYPRYLWHPCTINYTI 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

- Database : SPTREMBL_23.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phase.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	420	Q98UN4	Q98un4 hepatitis c
2	20	100.0	2436	Q81756	Q81756 hepatitis c
3	20	100.0	3011	Q91FES	Q91fes hepatitis c
4	20	100.0	3011	Q03463	Q03463 hepatitis c
5	20	100.0	3033	Q909A7	Q909a7 hepatitis c
6	19	95.0	3033	Q91ZA6	Q91za6 hepatitis c
7	18	90.0	206	Q81571	Q81571 hepatitis c
8	18	90.0	778	Q04184	Q04184 hepatitis c
9	18	90.0	3011	Q03579	Q03579 hepatitis c
10	18	90.0	3011	Q03610	Q03610 hepatitis c
11	18	90.0	3011	Q03610	Q03610 hepatitis c
12	18	90.0	3011	Q03610	Q03610 hepatitis c
13	18	90.0	3011	Q03609	Q03609 hepatitis c
14	17	85.0	376	Q81758	Q81758 hepatitis c
15	17	85.0	490	Q9PXN0	Q9pxn0 hepatitis c
16	17	85.0	3010	Q99AU2	Q99au2 hepatitis c

17	17	85.0	3010	12	P89966	P89966 hepatitis c
18	16	80.0	778	12	Q04185	Q04185 hepatitis c
19	15	75.0	128	12	Q8JYQ3	Q8jyq3 hepatitis c
20	15	75.0	128	12	Q8JYQ2	Q8jyq2 hepatitis c
21	15	75.0	128	12	Q8JYQ4	Q8jyq4 hepatitis c
22	15	75.0	129	12	Q8JYQ1	Q8jyq1 hepatitis c
23	15	75.0	143	12	Q81312	Q81312 hepatitis c
24	15	75.0	241	12	Q68448	Q68448 hepatitis c
25	15	75.0	241	12	Q68434	Q68434 hepatitis c
26	15	75.0	241	12	Q68435	Q68435 hepatitis c
27	15	75.0	241	12	Q68447	Q68447 hepatitis c
28	15	75.0	241	12	Q68454	Q68454 hepatitis c
29	15	75.0	364	12	Q37937	Q37937 hepatitis c
30	15	75.0	490	12	Q9PXN2	Q9pxn2 hepatitis c
31	15	75.0	625	12	Q9XPX3	Q9xp3 hepatitis c
32	15	75.0	640	12	Q68966	Q68966 hepatitis c
33	15	75.0	746	12	Q8JPM2	Q8jpm2 hepatitis c
34	15	75.0	1117	12	Q8QP91	Q8qp91 hepatitis c
35	15	75.0	2864	12	Q92973	Q92973 hepatitis c
36	15	75.0	2864	12	Q9WLK9	Q9wlk9 hepatitis c
37	15	75.0	2864	12	Q9WLL0	Q9wll0 hepatitis c
38	15	75.0	2864	12	Q9WLK8	Q9wlk8 hepatitis c
39	15	75.0	2864	12	Q92975	Q92975 hepatitis c
40	15	75.0	2864	12	Q92974	Q92974 hepatitis c
41	15	75.0	2864	12	Q92976	Q92976 hepatitis c
42	15	75.0	3010	12	Q9QIX1	Q9qix1 hepatitis c
43	15	75.0	3010	12	Q9J3G1	Q9j3g1 hepatitis c
44	15	75.0	3010	12	Q02829	Q02829 hepatitis c
45	15	75.0	3010	12	Q9J3G8	Q9j3g8 hepatitis c

ALIGNMENTS

RESULT 1

Q98UN4	PRELIMINARY;	PRT;	420 AA.
ID	Q98UN4		
AC	Q98UN4;		
DT	01-JUN-2001 (Tremblrel. 17, Created)		
DT	01-JUN-2001 (Tremblrel. 17, Last sequence update)		
DT	01-OCT-2002 (Tremblrel. 22, Last annotation update)		
DE	Genome polypeptide (Fragment).		
OS	Hepatitis C virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
OC	Hepacivirus.		
OX	NCBI_TaxID=11103;		
RN	[1]_TaxID=11103;		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-Q1a;		
RX	MEDLINE=20499063; PubMed=11044085;		
RA	Hadlock K.G., Lanford R.E., Perkins S., Rowe J., Yang Q., Levy S.,		
RA	Pileri P., Abrignani S., Fong S.K.,		
RT	"Human monoclonal antibodies that inhibit binding of hepatitis C virus		
RT	E2 protein to CD81 and recognize conserved conformational epitopes.";		
RL	J. Virol. 74:10407-10416(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-Q1a;		
RA	Keck Z.Y., Hadlock K.G., Yang Q.;		
RL	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF348704; AAK32685.1;		
DR	InterPro; IPR002519; HCV env.		
DR	InterPro; IPR002531; HCV NS1.		
DR	InterPro; IPR002518; HCV NS2.		
DR	Pfam; PF01539; HCV env; 1.		
DR	Pfam; PF01560; HCV NS1; 1.		
DR	Pfam; PF01538; HCV NS2; 1.		
DR	ProDom; PD186062; HCV NS1; 1.		
KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;		
KW	Polyprotein; Transmembrane.		
FT	NON_TER 1		
FT	CHAIN 1 >37 E1.		
FT	CHAIN 38 >400 E2.		

FT CHAIN 401 >420 P7.

FT NON_TER 420 420

SQ SEQUENCE 420 AA; 45982 MW; F28D0C3EE0CD081E CRC64;

Query Match 100.0%; Score 20; DB 12; Length 420;

Best Local Similarity 100.0%; Pred. No. 2.2e-15;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20

|||||

Db 261 CLVDYPYRLWHYPCTINYTI 280

RESULT 2

ID Q81756 PRELIMINARY; PRT; 2436 AA.

AC Q81756;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DT Genome polyprotein (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11103;

RN [1]

RP SEQUENCE FROM N.A.

RA Choo Q.-L., Richman K., Han J.;

RT "The nucleotide sequence of the Hepatitis C viral genome.";

RL Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.

DR EMBL; M32084; AAA45677.1; -.

DR HSSP; P27958; IALV.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002531; HCV NS1.

DR InterPro; IPR002518; HCV NS2.

DR InterPro; IPR004109; HCV NS3.

DR InterPro; IPR000745; HCV NS4.

DR InterPro; IPR001490; HCV NS4b.

DR InterPro; IPR002868; HCV NS5a.

DR InterPro; IPR002166; HCV RdRp.

DR InterPro; IPR001650; Helicase C.

DR InterPro; IPR007095; RNA_pol_DS_PS.

DR InterPro; IPR007094; RNA_pol_PSVir.

DR Pfam; PF01560; HCV NS1; 1.

DR Pfam; PF01538; HCV NS2; 1.

DR Pfam; PF02907; HCV NS3; 1.

DR Pfam; PF01006; HCV NS4a; 1.

DR Pfam; PF01001; HCV NS4b; 1.

DR Pfam; PF01506; HCV NS5a; 1.

DR Pfam; PF00271; helicase C; 1.

DR Pfam; PF00998; Viral RdRp; 1.

DR ProDom; PD186062; HCV NS1; 1.

DR SMART; SM00487; DEXDC; 1.

DR PROSITE; PS50507; RDRP_POSITIVE; 1.

DR PROSITE; PS50521; RDRP_VIRAL; 1.

KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;

KW Hydrolyase; Nonstructural protein; Polyprotein;

KW RNA-directed RNA polymerase; Transferase; Transmembrane.

FT NON_TER 1

FT NON_TER 2436 2436

SQ SEQUENCE 2436 AA; 264734 MW; D7B9872900BE3125 CRC64;

Query Match

Best Local Similarity 100.0%; Score 20; DB 12; Length 2436;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20

|||||

Db 157 CLVDYPYRLWHYPCTINYTI 176

RESULT 3.

Q91FE5

ID Q91FE5

AC Q91FE5; PRELIMINARY; PRT; 3011 AA.

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DT Genome polyprotein.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11103;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-21262212; PubMed-11369872;

RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;

RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype

sequence.";

RL J. Gen. Virol. 82:1291-1297 (2001).

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND MRNA (BY SIMILARITY).

DR EMBL; AF271632; RAF81759.1; -.

DR HSSP; P27958; IALV.

DR InterPro; IPR000345; CytC_heme_bind.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV_capsid.

DR InterPro; IPR002521; HCV core.

DR InterPro; IPR002519; HCV env.

DR InterPro; IPR002531; HCV NS1.

DR InterPro; IPR002518; HCV NS2.

DR InterPro; IPR004109; HCV NS3.

DR InterPro; IPR000745; HCV NS4b.

DR InterPro; IPR001490; HCV NS4a.

DR InterPro; IPR002868; HCV NS5a.

DR InterPro; IPR002166; HCV RdRp.

DR InterPro; IPR001650; Helicase C.

DR InterPro; IPR007095; RNA_pol_DS_PS.

DR InterPro; IPR007094; RNA_pol_PSVir.

DR Pfam; PF01543; HCV capsid; 1.

DR Pfam; PF01542; HCV core; 1.

DR Pfam; PF01539; HCV env; 1.

DR Pfam; PF01560; HCV NS1; 1.

DR Pfam; PF01538; HCV NS2; 1.

DR Pfam; PF02907; HCV NS3; 1.

DR Pfam; PF01006; HCV NS4a; 1.

DR Pfam; PF01001; HCV NS4b; 1.

DR Pfam; PF01506; HCV NS5a; 1.

DR Pfam; PF00271; helicase C; 1.

DR Pfam; PF00998; Viral RdRp; 1.

DR ProDom; PD186062; HCV NS1; 1.

DR SMART; SM00487; DEXDC; 1.

DR PROSITE; PS00190; CYTOCHROME C; 1.

DR PROSITE; PS50507; RDRP_POSITIVE; 1.

DR PROSITE; PS50521; RDRP_VIRAL; 1.

KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;

KW Hydrolyase; Nonstructural protein; Polyprotein;

KW RNA-directed RNA polymerase; Transferase; Transmembrane.

SQ SEQUENCE 3011 AA; 327124 MW; 2489CE74AC864E58 CRC64;

Query Match

Best Local Similarity 100.0%; Score 20; DB 12; Length 3011;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20

|||||

Db 607 CLVDYPYRLWHYPCTINYTI 626

RESULT 4

Q03463

ID Q03463

AC Q03463; PRELIMINARY; PRT; 3011 AA.

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DR ProDom; PD186062; HCV_NSI; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05021; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3033 AA; 329226 MW; 21492388CA05D8C CRC64;

Query Match 100.0%; Score 20; DB 12; Length 3033;
Best Local Similarity 100.0%; Pred. No. 1.1e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 20
DB 611 CLVDYPYRLWHYPCTINYTI 630
|||||

RESULT 6
Q91ZA6 PRELIMINARY; PRT; 3033 AA.
ID Q91ZA6
AC Q91ZA6
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD2a-1;
RA Itakura J., Nagayama K., Enomoto N., Kurosaki M., Watanabe H.,
RA Sato C.;
RT "Full length cDNA sequence of HCV genotype 2a.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: A
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF238481; AAF59940.1; -
DR HSP; P27958; 1A1V.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NSI.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01538; HCV_NSI; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05021; RDRP_VIRAL; 1.

KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3033 AA; 329444 MW; 9D0F544E06CA2B28 CRC64;

Query Match 95.0%; Score 19; DB 12; Length 3033;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTINYTI 19
DB 611 CLVDYPYRLWHYPCTINYTI 629
|||||

RESULT 7
Q81571 PRELIMINARY; PRT; 206 AA.
ID Q81571
AC Q81571
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE NS2 protein (Genome polyprotein) (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Hutchinson;
RX MEDLINE=91013116; PubMed=2170712;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA Yoshizawa H.;
RT "The 5'-terminal sequence of the hepatitis C virus genome.";
RL Jpn. J. Exp. Med. 60:167-177(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Hutchinson;
RA Inchauspe G., Zebedee S.L., Nasoff M.S., Sugitani M., Abe K.,
RA Prince A.M.;
RT "Cloning and nucleotide sequence analysis of structural and
RT nonstructural regions of the hutchinson strain of hepatitis C.";
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Hutchinson;
RX MEDLINE=89222455; PubMed=2496467;
RA Kuo G., Choo Q.-L., Alter H.J., Gitnick G.L., Redeker A.G.,
RA Purcell R.H., Miyamura T., Dienstag J.L., Alter M.J., Stevens C.E.,
RA Tegtmeyer G.E., Bonino F., Colombo M., Lee W.-S., Kuo C., Berger K.,
RA Shuster J.R., Overby L.R., Bradley D.W., Houghton M.;
RT "An assay for circulating antibodies to a major etiologic virus of
RT human non-A, non-B hepatitis.";
RL Science 244:362-365(1992).
DR EMBL; M55971; AAA45615.1; -
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 206
SQ SEQUENCE 206 AA; 22807 MW; 7FC6174D92325295 CRC64;

Query Match 90.0%; Score 18; DB 12; Length 206;
Best Local Similarity 100.0%; Pred. No. 2.6e-13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VDYPYRLWHYPCTINYTI 20
DB 119 VDYPYRLWHYPCTINYTI 136
|||||

RESULT 8

Q04184
ID Q04184 PRELIMINARY; PRT; 778 AA.
AC Q04184; Q81810;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_SEQUENCE FROM N.A.
RP MEDLINE=91195357; PubMed=1849654;
RX Ogata N., Alter H.J., Miller R.H., Purcell R.H.;
RA "Nucleotide sequence and mutation rate of the H strain of hepatitis C virus."
RT Proc. Natl. Acad. Sci. U.S.A. 88:3392-3396(1991).
RL -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: M62381; AAB02127.1; -;
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR ProDom: PD186062; HCV_NS2; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase; KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferrase; Transmembrane. SQ SEQUENCE 778 AA; 85143 MW; F7709172CD03E39B CRC64;
FT NON_TER 778
FT CHAIN 514 >778 STRUCTURAL PROTEIN.
Query Match 90.0%; Score 18; DB 12; Length 778;
Best Local Similarity 100.0%; Pred. No. 7.8e-13; Indels 0; Caps 0;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Caps 0;
QY 3 VDYPYRLWHYPTINYTI 20
DB 609 VDYPYRLWHYPTINYTI 626
RESULT 9
ID Q36579 PRELIMINARY; PRT; 3011 AA.
AC Q36579;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_SEQUENCE FROM N.A.
RP STRAIN=H77;
RX MEDLINE=97373636; PubMed=9228008;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M., Rice C.M.;
RT "Transmission of hepatitis C by intrahepatic inoculation with transcribed RNA."
RL Science 277:570-574(1997).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC EMBL: AF009606; AAB66324.1; -;
DR HSSP; P27958; 1HEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR00745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR001650; Helicase C.
DR InterPro: IPR007095; RNA pol DS PS.
DR InterPro: IPR007094; RNA pol PSvir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase; KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferrase; Transmembrane. SQ SEQUENCE 3011 AA; 327182 MW; E2E0EE809C63C1B9 CRC64;
Query Match 90.0%; Score 18; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.4e-12; Matches 18; Conservative 0; Mismatches 0; Indels 0; Caps 0;
QY 3 VDYPYRLWHYPTINYTI 20
DB 609 VDYPYRLWHYPTINYTI 626
RESULT 10
ID Q36610 PRELIMINARY; PRT; 3011 AA.
AC Q36610;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus strain H77.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=63746;
RN [1]_SEQUENCE FROM N.A.
RP STRAIN=H77;
RX MEDLINE=97385173; PubMed=9238047;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus are infectious when directly transfected into the liver of a chimpanzee."
RT Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: AF011753; AAB67038.1; -;
DR HSSP; P27958; 1HEI.

```

DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00507; RDRP_POSITIVE; 1.
DR PROSITE; PS05021; RDRP_VIRAL; 1.
DR PROSITE; PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydroxylase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327220 MW; 293F91430A0D4067 CRC64;

Query Match 90.0%; Score 18; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.4e-12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VDYPYRLWHYPCTINVTI 20
DB 609 VDYPYRLWHYPCTINVTI 626

RESULT 11
Q9ELS8 PRELIMINARY; PRT; 3011 AA.
AC Q9ELS8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=colonel;
RA Desai S.M., Devare S., Yamaguchi J.;
RT "Hepatitis C Virus.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF290978; AAGO2099.1; -.
DR HSSP; P27958; lHEI.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_env.
DR InterPro; IPR002519; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR002531; HCV_NS3.

```

```

DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
DR PROSITE; PS05037; RDRP_POSITIVE; 1.
DR PROSITE; PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydroxylase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327107 MW; A6BECF5A3B3EE13F CRC64;

Query Match 90.0%; Score 18; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.4e-12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VDYPYRLWHYPCTINVTI 20
DB 609 VDYPYRLWHYPCTINVTI 626

RESULT 12
O36609 PRELIMINARY; PRT; 3011 AA.
AC O36609;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus strain H77.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=63746;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97385173; PubMed=9238047;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus
RT are infectious when directly transfected into the liver of a
RT chimpanzee.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF011752; AAB67037.1; -.
DR HSSP; P27958; lHEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.

```

DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase.C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01539; HCV_core; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase.C; 1.
DR Pfam; PF00271; Helicase.C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05052; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydroxylase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327261 MW; 10D1C9702CA9B5DC CRC64;

Query Match 90.0%; Score 18; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.4e-12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VDYPYRLWHYPCTINYTI 20
|||||
DB 609 VDYPYRLWHYPCTINYTI 626

RESULT 13
ID O36608 PRELIMINARY; PRT; 3011 AA.
AC O36608;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-WAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus strain H77.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=63746;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97385173; PubMed=9238047;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus
are infectious when directly transfected into the liver of a
chimpanzee."
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743 (1997).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AF011751; AAB67036.1; --
DR HSP; P27958; 1HEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.

DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase.C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase.C; 1.
DR Pfam; PF00271; Helicase.C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05052; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydroxylase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327112 MW; 0B75E6B81CB5C198 CRC64;

Query Match 90.0%; Score 18; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.4e-12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VDYPYRLWHYPCTINYTI 20
|||||
DB 609 VDYPYRLWHYPCTINYTI 626

RESULT 14
ID Q81758 PRELIMINARY; PRT; 376 AA.
AC Q81758;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Genomic RNA (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91288501; PubMed=1648221;
RA Hijikata M., Kato N., Ootsuyama Y., Nakagawa M., Shimotohno K.;
RT "Gene mapping of the putative structural region of the hepatitis C
virus genome by in vitro processing analysis."
RL Proc. Natl. Acad. Sci. U.S.A. 88:5547-5551 (1991).
DR EMBL; D00757; BAA00653.1; --
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 376
SQ SEQUENCE 376 AA; 41562 MW; 1873339E70B81FDB CRC64;

Query Match 85.0%; Score 17; DB 12; Length 376;
Best Local Similarity 100.0%; Pred. No. 6.2e-12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTIN 17
|||||
DB 243 CLVDYPYRLWHYPCTIN 259

RESULT 15

Q9PXNO PRELIMINARY; PRT; 490 AA.
 ID Q9PXNO;
 AC Q9PXNO; (TREMELrel. 13, Created)
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
 DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
 DE Genome polyprotein [Contains: envelope glycoprotein E2/NS1 (GP68)]
 DE (Fragment)
 DE Hepatitis C virus type 2.
 OS Hepatitis C virus type 2.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC Hepacivirus.
 OX NCBI_TaxID=40271;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94351179; PubMed=7520922;
 RA Nakazawa T., Kato N., Ohkoshi S., Shibuya A., Shimotohno K.;
 RL J. Hepatol. 20:623-629(1994).
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 SQ SEQUENCE 490 AA; 53622 MW; D760B6A7182B14A6 CRC64;

Query Match 85.0%; Score 17; DB 12; Length 490;
 Best Local Similarity 100.0%; Pred. No. 7.7e-12;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVDYPYRLWHYPCTIN 17
 Db 446 CLVDYPYRLWHYPCTIN 462

Search completed: November 21, 2003, 21:08:21
 Job time : 24.45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:22 ; Search time 32.1 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-89

Perfect score: 20

Sequence: 1 PCTINYTFKIRMYGVGEH 20

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158728573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	16	AA884494
2	20	100.0	20	17	AA91012
3	20	100.0	20	23	AAO18718
4	20	100.0	121	14	AA832418
5	20	100.0	179	21	AA818529
6	20	100.0	192	19	AAW67009
7	20	100.0	254	22	AA868043
8	20	100.0	363	22	AA868042
9	20	100.0	402	14	AA834438

10	20	100.0	454	10	AA90183	Sequence of hepati
11	20	100.0	454	10	AA92049	Sequence encoded b
12	20	100.0	454	21	AA818526	Protein encoded by
13	20	100.0	480	21	AA833992	HCV-1 E2/NS1 prote
14	20	100.0	531	22	AAE02622	Chimeric HCV E2661
15	20	100.0	621	14	AA833185	Sequence of subfra
16	20	100.0	622	14	AA833591	HCV CKS-NS1-NS1S
17	20	100.0	622	22	AA869010	HCV recombinant an
18	20	100.0	637	24	AA857410	Hepatitis C virus
19	20	100.0	663	17	AA892935	HCV1 E2 + NS2 poly
20	20	100.0	663	20	AAW67615	Hepatitis C virus
21	20	100.0	738	14	AA833592	HCV CKS-full lengt
22	20	100.0	738	22	AA869011	HCV recombinant an
23	20	100.0	2435	13	AA825135	HCV polypeptide 1.
24	20	100.0	2436	10	AA892050	Sequence encoded i
25	20	100.0	2436	10	AA892088	Peptide encoded by
26	20	100.0	2436	13	AA828582	HCV amino acid seq
27	20	100.0	2772	11	AA808123	Hepatitis C virus
28	20	100.0	2772	21	AA818540	Protein encoded by
29	20	100.0	2894	13	AA824440	Composite HCV HC-J
30	20	100.0	2894	16	AA870230	Composite hepatiti
31	20	100.0	2955	11	AA808124	Hepatitis C virus
32	20	100.0	2955	20	AAV14975	Amino acid sequenc
33	20	100.0	2955	21	AA818541	Polyprotein encode
34	20	100.0	3011	13	AA821519	Compiled HCV sequ
35	20	100.0	3011	14	AA831621	Hepatitis C virus
36	20	100.0	3011	17	AA890931	Hepatitis C virus
37	20	100.0	3011	18	AAW34480	HCV polyprotein.
38	20	100.0	3011	19	AAW40038	Hepatitis C virus
39	20	100.0	3011	23	AAE22049	Hepatitis C virus
40	20	100.0	3011	23	AAE22052	Hepatitis C virus
41	16	80.0	737	14	AA838281	NAHB hepatitis vir
42	16	80.0	3015	22	AA830730	Amino acid sequenc
43	16	80.0	3015	22	AA830731	Amino acid sequenc
44	16	80.0	3015	22	AA830732	Amino acid sequenc
45	16	80.0	3015	22	AA830733	Amino acid sequenc

ALIGNMENTS

```
RESULT 1
AA884494
ID AA884494 standard; peptide; 20 AA.
XX
AC AA884494;
XX
DT 06-JAN-1997 (first entry)
XX
DE Hepatitis C virus T-cell epitope peptide NS1-29 (residues 619-638).
XX
KW Hepatitis C virus; HCV; immunogen; non-structural region;
KW immunodominant; T cell epitope; vaccine.
XX
OS Hepatitis C virus.
XX
PN WO9512677-A2.
XX
PD 11-MAY-1995.
XX
PF 28-OCT-1994; 94WO-EF03555.
XX
PR 04-NOV-1993; 93EP-0402718.
XX
(PA (INNO-) INNOGENETICS NV.
XX
PI Deleys R, Leroux-Roels G, Maertens G;
XX
DR WPI; 1995-193822/25.
XX
PT Hepatitis C Virus immunogenic polypeptide contg. a T-cell
PT stimulating epitope - from core, E1, E2 and NS3 regions, useful in
PT production of vaccines, therapeutic agents, etc.
```

XX PS Claim 26; Page 71; 105pp; English.

XX CC The present sequence is a specifically claimed example of a

CC CC T-cell epitope-containing peptide derived from hepatitis C virus.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 16; Length 20;

Best Local Similarity 100.0%; Pred. No. 5.4e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINTYTFKIRMYGVGVH 20

DB 1 PCTINTYTFKIRMYGVGVH 20

RESULT 2

AAR91012

ID AAR91012 standard; peptide; 20 AA.

XX AC AAR91012;

XX DT 25-SEP-1996 (first entry)

XX DE HCV E2 peptide E2-29 for competition studies.

XX KW HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;

KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.

XX OS Synthetic.

XX PN WO9604385-A2.

XX PD 15-FEB-1996.

XX PF 31-JUL-1995; 95WO-EP03031.

XX PR 29-JUL-1994; 94EP-0870132.

XX PA (INNO-) INNOGENETICS NV.

XX PI Bosman F, Buyse M, De Martynoff G, Maertens G;

XX PS WPI; 1996-129401/13.

XX PT Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope

PT proteins - in presence of disulphide bond cleavage agent, to

PT produce proteins suitable for direct use in vaccines or diagnostic

PT assays of HCV

XX PS Example 7; Page 67; 146pp; English.

XX CC AAR90965-R90980 and AAR90995-R91015 represent synthetic hepatitis C

CC virus (HCV) E1 and E2 peptides used in competition studies. This

CC sequence represents a synthetic E2 peptide, and corresponds to residues

CC 619-638 of the E2 protein sequence. These sequences are useful for in

CC vitro monitoring of HCV disease, or prognosis of the response to

CC interferon treatment of patients suffering from HCV infection. These

CC sequences compete with the proteins produced by AAT12704-T12709 and

CC AAT12961-T12974, which are included in vectors for the production of

CC recombinant E1, E2, and E1/E2. The recombinant proteins can then be

CC isolated and purified by carrying out a disulphide bond cleavage, or a

CC reduction step with a disulphide bond cleavage agent, after lysis of

CC recombinant host cells. The constructs containing the purified HCV

CC envelope proteins can be used for vaccinating humans against HCV, for in

CC vitro detection of HCV antibodies in a sample, and in a serotyping assay

CC for detecting one or more serological types of HCV present in a

CC biological sample. The constructs can also be immobilised on a solid

CC substrate and incorporated into a reversed phase hybridisation assay for

CC determining the presence or the genotype of HCV. The new purification

CC method preserves the conformation of the recombinantly expressed E1, E2

CC and E1/E2, and eliminates contaminating proteins. Antigens isolated

XX CC using this method are more reactive with human sera than those isolated

CC by known techniques.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 17; Length 20;

Best Local Similarity 100.0%; Pred. No. 5.4e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINTYTFKIRMYGVGVH 20

DB 1 PCTINTYTFKIRMYGVGVH 20

RESULT 3

AAO18718

ID AAO18718 standard; Peptide; 20 AA.

XX AC AAO18718;

XX DT 24-OCT-2002 (first entry).

XX DE Hepatitis C virus E2 protein derived peptide E2-29.

XX KW Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide;

KW immunostimulant; vaccine.

XX OS Hepatitis C virus.

XX PN WO200255548-A2.

XX PD 18-JUL-2002.

XX PF 11-JAN-2002; 2002WO-EP00219.

XX PR 11-JAN-2001; 2001US-260699P.

XX PR 30-AUG-2001; 2001US-315768P.

XX PA (INNO-) INNOGENETICS NV.

XX PI Maertens G, Bosman F, Buyse M;

XX PS WPI; 2002-599657/64.

XX PT New therapeutic vaccine compositions comprising at least one purified

PT recombinant hepatitis C virus (HCV) single or specific oligomeric

PT recombinant envelope protein E1 or E2, useful for immunizing humans

PT from HCV infection

XX PS Example 7; Page 231; 243pp; English.

XX CC The present invention relates to new therapeutic vaccine compositions for

CC inducing hepatitis C virus (HCV)-specific antibodies, comprising a

CC composition containing at least one purified recombinant HCV single or

CC specific oligomeric recombinant envelope proteins selected from an E1 and

CC an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are

CC useful for inducing HCV-specific antibodies or for immunising humans

CC against HCV. The recombinant HCV E1 and/or E2 proteins are useful as

CC vaccines or therapeutics, in HCV screening and confirmatory antibody

CC tests, for raising antibodies, in the preparation of medicament, and for

CC in vitro monitoring of HCV disease or prognosis of the response to

CC treatment of patients suffering from HCV infection. The present sequence

CC is a peptide derived from the proteins of the invention.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 23; Length 20;

Best Local Similarity 100.0%; Pred. No. 5.4e-14;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINTYTFKIRMYGVGVH 20

DB 1 PCTINTYTFKIRMYGVGVH 20

RESULT 4
 AAR32418
 ID AAR32418 standard; peptide; 121 AA.
 XX AC
 XX AAR32418;
 DT 25-MAR-2003 (updated)
 DT 03-JUL-1993 (first entry)
 XX
 DE Sequence of immunogenic domain of E2/NS1 region of hepatitis C virus
 DE (HCV) genome encompassing AAs 600-720.
 XX
 KW Immunogenic peptide; hepatitis C virus; immunogenic domain;
 KW monoclonal antibody; diagnosis; detection; therapy.
 XX
 XX Synthetic.
 OS
 XX WO9304205-A1.
 PN
 XX
 PD 04-MAR-1993.
 XX
 PF 21-AUG-1992; 92WO-US07189.
 XX
 PR 21-AUG-1991; 91US-0748292.
 XX
 PA (ABBO) ABBOTT LAB.
 XX
 PI Dailey SH, Deesai SM, Devare SG, Johnson JE, Mehta SU;
 XX
 DR WPI; 1993-094030/11.
 XX
 PT Monoclonal antibodies specific for hepatitis C virus E2-NS1
 PT antigen - useful for diagnosis and evaluation of HCV infections
 PT and in differentiation studies
 XX
 PS Example; Page 29; 48pp; English.
 XX
 CC Immunogenic domains of E2/NS1 region of HCV genome encompassing AAs
 CC 600-720 were mapped with PEPSKAN analysis. FAB dimers of IgG
 CC purified from sera of individuals seropositive for antibodies to HCV
 CC proteins were used as the primary antibody for the serological
 CC analysis. Based on the reactivity these sera in EIA, four AA
 CC sequences (AAR32419,R32420,AAR33182,R33183) were identified as
 CC immunogenic domains. Each of these four sequences and an additional
 CC sequence, which was the combination of the two most immunogenic
 CC sequences (AAR33184) were synthesised. Based on the EIA reactivity of
 CC a panel of HCV positive sera, peptide AAR33184 was chosen as the
 CC immunogen for the generation of monoclonal antibodies to HCV NS1.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 121 AA;
 Query Match 100.0%; Score 20; DB 14; Length 121;
 Best Local Similarity 100.0%; Pred. No. 2.4e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PCTINTYTFKIRMYGGVEH 20
 DB 20 PCTINTYTFKIRMYGGVEH 39
 RESULT 5
 AAB18529
 ID AAB18529 standard; Protein; 179 AA.
 XX AC
 XX AAB18529;
 DT 15-JAN-2001 (first entry)
 DE Protein encoded by a novel hepatitis C virus cDNA clone 131.

KW Hepatitis C virus; HCV; antisense polynucleotide; polypeptide; viral infectivity; viral replication.
 XX
 OS Hepatitis C virus.
 XX
 PN EP1034785-A2.
 XX
 PD 13-SEP-2000.
 XX
 PF 16-MAR-1990; 2000EP-0109602.
 XX
 PR 17-MAR-1989; 89US-0325338.
 PR 20-APR-1989; 89US-0341334.
 PR 18-MAY-1989; 89US-0355002.
 PR 16-MAR-1990; 90EP-0302866.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Houghton M, Choo Q, Kuo G;
 XX
 DR WPI; 2000-566891/53.
 DR N-PSDB; AAA75285.
 XX
 PT Novel composition comprising a hepatitis C virus antisense
 PT polynucleotide which is complementary to or corresponds to a sense
 PT strand of the virus genome, and selectively hybridises to it -
 XX
 PS Example; Fig 5; 75pp; English.
 XX
 CC The specification describes a pharmaceutical composition which
 CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
 CC HCV is characterized by a positive stranded RNA genome which has
 CC 40% homology at the polypeptide level to a HCV polypeptide. The
 CC antisense polynucleotide binds to cellular polynucleotides which
 CC enhance and/or are required for viral infectivity, replicative
 CC ability or chronicity. The antisense polynucleotides may also be
 CC designed to bind with high specificity, to be of increased stability,
 CC to be stable and to have low toxicity. The composition also comprises
 CC an agent which causes viral RNA to be inactive. The composition
 CC is used for preventing HCV replication in a system. The present
 CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
 CC course of the invention.
 XX
 SQ Sequence 179 AA;
 Query Match 100.0%; Score 20; DB 21; Length 179;
 Best Local Similarity 100.0%; Pred. No. 3.3e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PCTINTYTFKIRMYGGVEH 20
 DB 109 PCTINTYTFKIRMYGGVEH 128
 RESULT 6
 AAW67009
 ID AAW67009 standard; protein; 192 AA.
 XX
 AC AAW67009;
 XX
 DT 02-MAR-1999 (first entry)
 XX
 DE HCV nucleocapsid core protein.
 XX
 KW Hepatitis C virus; HCV; nucleocapsid; core protein; analogue; antibody;
 KW non-structural protein; thioamide bond; peptide bond.
 XX
 OS Hepatitis C virus.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 79 /note= "given in specification as Try"
 XX

PN JP10226698-A.
 XX 25-AUG-1998.
 XX 19-FEB-1997; 97JP-0034702.
 XX 19-FEB-1997; 97JP-0034702.
 XX (KYOW) KYOWA MEDEX KK.
 XX WPI; 1998-515103/44.
 XX Determination of antibody in sample - uses peptide analog absorbed
 PT or chemically bound on carrier as antigen
 XX
 XX Disclosure; Page 4; 13pp; Japanese.
 XX This sequence represents the Hepatitis C virus (HCV) nucleocapsid core
 CC protein. The invention relates to peptide analogues derived from HCV
 CC proteins, e.g. AAW67417-W67426, which can be used for the determination
 CC of anti-HCV antibodies in a sample. Preferably the peptide analogues
 CC contain one or more thioamide peptide bonds where at least one oxygen
 CC atom of the peptide bond is replaced by sulphur atom. The peptide
 CC analogues can be adsorbed or chemically bound to a carrier.
 XX
 XX Sequence 192 AA;
 SQ
 Query Match 100.0%; Score 20; DB 19; Length 192;
 Best Local Similarity 100.0%; Pred. No. 3.5e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PCTINTYTFKIRMYVGVEH 20
 |||||
 DB 169 PCTINTYTFKIRMYVGVEH 188
 |||||
 RESULT 7
 AAB68043
 ID AAB68043 standard; protein; 254 AA.
 AC AAB68043;
 XX 29-JUN-2001 (first entry)
 DT
 DE Amino acid sequence of water soluble variant of envelope E2 protein.
 XX
 XX Eo protein; HCV; envelope protein; E2 protein; HCV infection;
 KW HCV attachment.
 KW Synthetic.
 XX Hepatitis C virus.
 OS
 PN WO200122984-A1.
 XX
 XX 05-APR-2001.
 PD
 XX 26-SEP-2000; 2000WO-US26395.
 PF
 XX 29-SEP-1999; 99US-0407430.
 PR
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 PA
 XX Worman HJ, Mamiya N;
 PI WPI; 2001-273486/28.
 XX
 XX Treating or preventing hepatitis C virus infection in a subject,
 PT involves administering hepatitis C virus envelope protein E2 binding
 PT agents -
 XX
 XX Claim 5; Fig 8; 46pp; English.
 PS
 XX The present sequence represents a water soluble variant of a Hepatitis C

CC virus (HCV) envelope E2 protein. Agents that bind to the HCV E2 protein
 CC (such as the human Eo protein), and so inhibit the attachment of HCV onto
 CC cells (especially liver cells), are used to treat HCV infections in
 CC mammals, in particular humans. The specification also describes a method
 CC for identifying a compound which can be used for treating or preventing
 CC HCV in a subject and which can inhibit the attachment of HCV onto cells
 CC by inhibiting the binding of HCV envelope E2 protein to a cellular
 CC protein associated with HCV attachment and entry into cells. The method
 CC comprises incubating the compound, HCV envelope E2 protein or its variant
 CC and a cellular protein capable of specifically binding to the HCV E2
 CC protein under suitable reaction conditions; determining the interactions
 CC between HCV envelope E2 protein and cellular protein in the presence and
 CC absence of the compound; and comparing the interaction to identify a
 CC compound which can inhibit the attachment of HCV onto cells.
 XX
 XX Sequence 254 AA;
 SQ
 Query Match 100.0%; Score 20; DB 22; Length 254;
 Best Local Similarity 100.0%; Pred. No. 4.4e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PCTINTYTFKIRMYVGVEH 20
 |||||
 DB 214 PCTINTYTFKIRMYVGVEH 233
 |||||
 RESULT 8
 AAB68042
 ID AAB68042 standard; protein; 363 AA.
 XX
 AC AAB68042;
 XX 29-JUN-2001 (first entry)
 DT
 DE Amino acid sequence of a Hepatitis C virus envelope E2 protein.
 XX
 XX Eo protein; HCV; envelope protein; E2 protein; HCV infection;
 KW HCV attachment.
 KW Hepatitis C virus.
 OS
 PN WO200122984-A1.
 XX
 XX 05-APR-2001.
 PD
 XX 26-SEP-2000; 2000WO-US26395.
 PF
 XX 29-SEP-1999; 99US-0407430.
 PR
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 PA
 XX Worman HJ, Mamiya N;
 PI WPI; 2001-273486/28.
 XX
 XX Treating or preventing hepatitis C virus infection in a subject,
 PT involves administering hepatitis C virus envelope protein E2 binding
 PT agents -
 XX
 XX Claim 3; Fig 7; 46pp; English.
 PS
 XX The present sequence represents a Hepatitis C virus (HCV) envelope
 CC E2 protein. Agents that bind to the HCV E2 protein (such as the human
 CC Eo protein), and so inhibit the attachment of HCV onto cells
 CC (especially liver cells), are used to treat HCV infections in mammals,
 CC in particular humans. The specification also describes a method for
 CC identifying a compound which can be used for treating or preventing
 CC HCV in a subject and which can inhibit the attachment of HCV onto cells
 CC by inhibiting the binding of HCV envelope E2 protein to a cellular
 CC protein associated with HCV attachment and entry into cells. The method
 CC comprises incubating the compound, HCV envelope E2 protein or its variant
 CC and a cellular protein capable of specifically binding to the HCV E2
 CC protein under suitable reaction conditions; determining the interactions

CC between HCV envelope E2 protein and cellular protein in the presence and
 CC absence of the compound; and comparing the interaction to identify a
 CC compound which can inhibit the attachment of HCV onto cells.

XX Sequence 363 AA;

Query Match 100.0%; Score 20; DB 22; Length 363;
 Best Local Similarity 100.0%; Pred. No. 5.9e-13;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTFKIRMYGVGEH 20

DB 236 PCTINYTFKIRMYGVGEH 255

RESULT 9

AAR34438

ID AAR34438 standard; Protein; 402 AA.

XX AAR34438;

XX 25-MAR-2003 (updated)

DT 09-AUG-1993 (first entry)

XX Sequence of glycoprotein E2/NS1 in clone HCV1.

DE Hepatitis C virus; envelope protein; glycoprotein; E2/NS1;

KW diagnostic reagent.

KW Hepatitis C virus.

OS Hepatitis C virus.

XX EP537626-A1.

PN 21-APR-1993.

XX 08-OCT-1992; 92EP-0117191.

XX 08-OCT-1991; 91JP-0260824.

XX (NAHE-) NAT INST OF HEALTH.

XX Harada S, Honda Y, Miyamura T, Saito I;

XX WPI; 1993-127516/16.

DR N-PSDB; AAQ40330.

XX Diagnostic reagent for hepatitis C virus - comprises second

PT envelope protein or first non-structural protein encoded by HCV

PT gene and has sugar chain

XX Claim 2; Pages 30-32; 58pp; English.

CC Glycoprotein E2/NS1 is derived from the second envelope protein or

CC first non-structural protein encoded by the genome of HCV. The

CC nucleic acid is extracted from the serum of the patient of hepatitis

CC C. The serum is pref. mixed with transfer RNA (tRNA) as a carrier

CC of virus RNA. As a technique of cloning cDNA from the nucleic acid,

CC it is preferred to use polymerase chain reaction method. In the

CC reaction, any commercially available random primers or synthesized

CC DNA having a base sequence similar to that of primer AS1 may be used

CC as a primer. Representative examples of sense primers include S1.

CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 402 AA;

Query Match 100.0%; Score 20; DB 14; Length 402;

Best Local Similarity 100.0%; Pred. No. 6.4e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTFKIRMYGVGEH 20

DB 280 PCTINYTFKIRMYGVGEH 299

RESULT 11

AAP92049

ID AAP92049 standard; protein; 454 AA.

XX AAP92049;

XX 25-MAR-2003 (updated)

DT 21-JAN-1991 (first entry)

XX Sequence encoded by segment of the hepatitis C virus (HCV) cDNA sequence

DE in clone K9-1.

XX

RESULT 10

AAP90183

ID AAP90183 standard; protein; 454 AA.

XX AAP90183;

XX 25-MAR-2003 (updated)

DT 01-NOV-1989 (first entry)

XX Sequence of hepatitis C virus cDNA insert in clone k9-1.

DE Hepatitis C virus; clone k9-1; probe; vaccine.

XX Pan troglodytes.

XX Key

FT Region

XX Location/Qualifiers

XX 97..454

XX GB2212511-A.

XX 26-JUL-1989.

XX 18-NOV-1988; 88GB-0027024.

XX 18-NOV-1987; 87US-0122714.

XX 30-DEC-1987; 87US-0139886.

XX 26-FEB-1988; 88US-0161072.

XX 26-OCT-1988; 88US-0263584.

XX (CHIR) CHIRON CORPORATION.

XX Houghton M, Choo QL, Kuo G;

XX WPI; 1989-215054/30.

XX N-PSDB; AAN90335.

XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,

XX polypeptide(s) and antibodies for diagnosis, prevention and

XX treatment of infection.

XX Disclosure; fig 46; 235pp; English.

XX The sequence is the peptide encoded by the hepatitis C virus

XX (HCV) cDNA insert in clone k9-1 (see AAN90335). The polypeptides

XX are used to diagnose HCV-induced NANBH, to raise antibodies for

XX immunoassay or treatment, or to produce vaccines.

XX The region shown overlaps the cDNA of AAN90327.

XX (Updated on 25-MAR-2003 to correct PR field.)

XX Sequence 454 AA;

Query Match 100.0%; Score 20; DB 10; Length 454;

Best Local Similarity 100.0%; Pred. No. 7.1e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTFKIRMYGVGEH 20

DB 169 PCTINYTFKIRMYGVGEH 188

KW Non-a non-B hepatitis; probe; vaccine; diagnosis;
 KW passive immunotherapy; antigen.

XX Hepatitis C virus.

PN EP318216-A.

XX 31-MAY-1989.

XX 18-NOV-1988; 88EP-0310922.

XX 18-NOV-1987; 87US-0122714.

PR 30-DEC-1987; 87US-0139886.

PR 26-FEB-1988; 88US-0161072.

PR 06-MAY-1988; 88US-0191263.

PR 26-OCT-1988; 88US-0263584.

PR 14-NOV-1988; 88US-0271450.

XX (CHIR) CHIRON CORP.

XX Houghton M, Choo QL, Kuo G;

DR WPI; 1989-159274/22.

DR N-PSDB; AAN92105.

XX Purified hepatitis C virus -

PT and assorted nucleic acids and polypeptide(s)

PT Example; Fig 46-1 - 46-2; 139pp; English.

XX Purified hepatitis C virus (HCV) and purified or recombinant HCV nucleic

CC acids (NAs), encoding HCV polynucleotides or epitopes, and polypeptides

CC are claimed. HCV is a causative agent of non-A, non-B hepatitis (NANBH).

CC The NAs may be used to design probes for detn. of HCV NAs in samples.

CC The polypeptides may be used as immunoassay reagents and vaccines, and

CC to produce antibodies useful for diagnosis and passive immunotherapy.

CC the purified virus may also be used in vaccines.

CC (Updated on 25-MAR-2003 to correct PR field.)

CC (Updated on 25-MAR-2003 to correct PA field.)

XX SQ Sequence 454 AA;

Query Match 100.0%; Score 20; DB 10; Length 454;

Best Local Similarity 100.0%; Pred. No. 7.1e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINTYTFKIRMYGVGVH 20

DB 169 PCTINTYTFKIRMYGVGVH 188

RESULT 12

AAB18526

ID AAB18526 standard; Protein; 454 AA.

XX AAB18526;

XX 15-JAN-2001 (first entry)

XX Protein encoded by a novel hepatitis C virus cDNA clone k9-1.

XX Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;

KW viral infectivity; viral replication.

XX Hepatitis C virus.

OS EP1034785-A2.

XX 13-SEP-2000.

XX 16-MAR-1990; 2000EP-0109602.

XX 17-MAR-1989; 89US-0325338.

PR 20-APR-1989; 89US-0341334.

PR 18-MAY-1989; 89US-0355002.

XX 16-MAR-1990; 90EP-0302866.

XX (CHIR) CHIRON CORP.

XX Houghton M, Choo Q, Kuo G;

XX WPI; 2000-566891/53.

XX N-PSDB; AAA75282.

XX Novel composition comprising a hepatitis C virus antisense

PT polynucleotide which is complementary to or corresponds to a sense

PT strand of the virus genome, and selectively hybridizes to it -

XX Example; Fig 2; 75pp; English.

XX The specification describes a pharmaceutical composition which

CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The

CC HCV is characterized by a positive stranded RNA genome which has

CC 40% homology at the polypeptide level to a HCV polyprotein. The

CC antisense polynucleotide binds to cellular polynucleotides which

CC enhance and/or are required for viral infectivity, replicative

CC ability or chronicity. The antisense polynucleotides may also be

CC designed to bind with high specificity, to be of increased stability,

CC to be stable and to have low toxicity. The composition also comprises

CC an agent which causes viral RNA to be inactive. The composition

CC is used for preventing HCV replication in a system. The present

CC sequence is encoded by a novel HCV cDNA sequence, which is used in the

CC course of the invention.

XX SQ Sequence 454 AA;

Query Match 100.0%; Score 20; DB 21; Length 454;

Best Local Similarity 100.0%; Pred. No. 7.1e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINTYTFKIRMYGVGVH 20

DB 169 PCTINTYTFKIRMYGVGVH 188

RESULT 13

AAR33992

ID AAR33992 standard; Protein; 480 AA.

XX AAR33992;

XX 25-MAR-2003 (updated)

XX 26-JUL-1993 (first entry)

XX HCV-1 E2/NS1 protein.

XX Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;

XX HCV; asymptomatic; chronically infected; epitope; viral isolate;

KW domain; immunological; cross-reactive; envelope protein; vaccine;

KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.

XX Synthetic.

XX WO9306126-A1.

XX 01-APR-1993.

XX 11-SEP-1992; 92WO-US07683.

XX 13-SEP-1991; 91US-0759575.

XX (CHIR) CHIRON CORP.

XX Houghton M, Weiner AJ;

XX WPI; 1993-117468/14.

XX Immuno-reactive hepatitis C virus polypeptide compans. - contg.
PT at least 2 sequences from the first variable domain of distinct
PT HCV isolates
XX
PS Disclosure; Fig 3; 106pp; English.
XX
XX The sequences given in AAR33992-002 represent a portion of the E2/NS1
CC protein encoded by group I and group II HCV isolates, from amino acid
CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
CC 30 amino acids which shows large variation between nearly all isolates.
CC This is an important immunoreactive domain. This putative envelope
CC glycoprotein E2/NS1 may correspond to the gp33(BVDV)/gp55 (hog cholera
CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
CC flaviviruses, both of which confer protective immunity in hosts
CC vaccinated with these polypeptides. It has been discovered that a
CC number of important HCV epitopes vary among viral isolates and that
CC these epitopes can be mapped to specific domains. This meant that
CC immunologically cross-reactive polypeptides which focus on variable
CC rather than constant domains can be produced. See also AAR339134-48
CC and AAR33982-91.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 480 AA;
SQ
Query Match 100.0%; Score 20; DB 14; Length 480;
Best Local Similarity 100.0%; Pred. No. 7.5e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PCTINYTIKIRMYGVGVH 20
DB 250 PCTINYTIKIRMYGVGVH 269
|||||
RESULT 14
AAR33992
ID AAE02622 standard; Protein; 531 AA.
XX
AC AAE02622;
XX
XX 06-AUG-2001 (first entry)
XX
DE Chimeric HCV E2661-HBsAg S domain encoded by pCMV-II-E2661-SAg.
XX
XX Virus-like particle; immunogen; hepatitis B virus surface antigen;
KW HBsAg; HBV; chimeric antigen; hepatitis C virus; HCV; virucide; vaccine;
KW plasmid pCMV-II-E2661-SAg; E2 envelope glycoprotein; S domain.
XX
OS Chimeric - Hepatitis C virus.
OS Chimeric - Hepatitis B virus.
XX
XX Key Location/Qualifiers
FH Region 1...303
FT /note= "HCV 661 E2 envelope glycoprotein"
FT Region 306..531
FT /note= "HBsAg S domain"
XX
XX WO200138358-A2.
XX
XX 31-MAY-2001.
PD
XX 22-NOV-2000; 2000WO-US32249.
PF
XX 24-NOV-1999; 99US-0167224.
XX
XX (CHIR) CHIRON CORP.
XX
XX Selby M, Glazer E, Houghton M;
XX
XX WPI: 2001-367661/38.
DR N-PSDB; AAD06793.
XX

PT Virus-like particle for use as an immunogen, comprising a first
PT hepatitis B virus surface antigen (HBsAg) and chimeric antigen
PT comprising a second HBsAg covalently linked to hepatitis C immunogenic
PT polypeptide -
XX
PS Claim 29; Fig 4; 115pp; English.
XX
XX The invention relates to a virus-like particle for use as an immunogen,
CC comprising a first hepatitis B virus surface antigen (HBsAg) and a
CC chimeric antigen comprising a second HBsAg which is covalently linked to
CC an hepatitis C virus (HCV) immunogenic polypeptide, where the first and
CC the second HBsAg each comprise a substantially complete S domain.
CC The virus-like particle is useful as immunogen and as vaccine.
CC The present sequence is a chimeric antigen comprising HCV 661 E2
CC envelope glycoprotein and HBsAg S domain encoded by plasmid
CC pCMV-II-E2661-SAg.
XX
XX Sequence 531 AA;
SQ
Query Match 100.0%; Score 20; DB 22; Length 531;
Best Local Similarity 100.0%; Pred. No. 8.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PCTINYTIKIRMYGVGVH 20
DB 261 PCTINYTIKIRMYGVGVH 280
|||||
RESULT 15
AAR33185
ID AAR33185 standard; peptide; 621 AA.
XX
AC AAR33185;
XX
XX 25-MAR-2003 (updated)
DT 03-JUL-1993 (first entry)
XX
DE Sequence of subfragment pHCV80 (AAs 365-731) of the hepatitis
DE C virus (HCV) genome.
XX
KW Immunogenic peptide; hepatitis C virus; immunogenic domain;
KW monoclonal antibody; diagnosis; detection; therapy.
XX
OS Synthetic.
XX
XX WO9304205-A1.
XX
XX 04-MAR-1993.
PD
XX 21-AUG-1992; 92WO-US07189.
PF
XX 21-AUG-1991; 91US-0748292.
PR
XX (ABBO) ABBOTT LAB.
XX
XX Dailey SH, Desai SM, Devare SG, Johnson JE, Mehta SU;
PI WPI: 1993-094030/11.
XX
XX Monoclonal antibodies specific for hepatitis C virus E2-NS1
PT antigen - useful for diagnosis and evaluation of HCV infections
PT and in differentiation studies
XX
XX Example; Pages 31-34; 48pp; English.
XX
XX Immunogenic domains of E2/NS1 region of HCV genome encompassing AAs
CC 600-720 were mapped with PEPSCAN analysis. Based on the EIA reactivity
CC of a panel of HCV positive sera, peptide AAR33184 was chosen as the
CC immunogen for the generation of monoclonal antibodies to HCV NS1.
CC Several individual oligos representing AAs 165-731 of HCV genome
CC were ligated and cloned as three separate EcoRI/BamHI subfragments
CC into the CKS fusion vector pJ0200. The sequences of these three
CC subfragments - pHCV80, pHCV77 and pHCV65 are in AAR33185-7.

Query Match 100.0%; Score 20; DB 14; Length 621;
Best Local Similarity 100.0%; Pred. No. 9.2e-13;
Matches 20: Conservative 0; Mismatches 0; Indels

509 PCTINYTI FKIRMYVGGVEH 528

Search completed: November 21, 2003, 20:58:07
Job time : 33.15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.6 Seconds
(without alignments)
79.832 Million cell updates/sec

Title: US-09-973-025-89

Perfect score: 20

Sequence: 1 PCTINYTFKIRMYGVGEH 20

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/1/aaa/5A COMB.pep:*
- 2: /cgn2_6/prodata/1/aaa/5B COMB.pep:*
- 3: /cgn2_6/prodata/1/aaa/6A COMB.pep:*
- 4: /cgn2_6/prodata/1/aaa/6B COMB.pep:*
- 5: /cgn2_6/prodata/1/aaa/PCTUS COMB.pep:*
- 6: /cgn2_6/prodata/1/aaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-08-612-973-89
2	20	100.0	20	3	US-08-927-597-89
3	20	100.0	20	4	US-08-635-886C-45
4	20	100.0	68	4	US-08-635-886C-250
5	20	100.0	88	1	US-08-440-103-26
6	20	100.0	88	1	US-08-440-103-27
7	20	100.0	88	1	US-08-440-542-26
8	20	100.0	88	1	US-08-440-542-27
9	20	100.0	88	1	US-08-231-368-26
10	20	100.0	88	1	US-08-231-368-27
11	20	100.0	88	1	US-08-440-210-26
12	20	100.0	88	1	US-08-440-210-27
13	20	100.0	88	4	US-09-046-604-26
14	20	100.0	88	4	US-09-046-604-27
15	20	100.0	121	1	US-07-748-292-1
16	20	100.0	179	3	US-08-444-818-77
17	20	100.0	402	1	US-08-460-806-13
18	20	100.0	402	1	US-08-325-630-13
19	20	100.0	403	2	US-08-483-695-39
20	20	100.0	403	2	US-07-965-285-39
21	20	100.0	403	2	US-08-487-231-39
22	20	100.0	403	3	US-09-201-912-39
23	20	100.0	414	1	US-07-748-292-8
24	20	100.0	454	3	US-08-444-818-73
25	20	100.0	480	1	US-08-440-103-14
26	20	100.0	480	1	US-08-440-542-14
27	20	100.0	480	1	US-08-231-368-14

28 20 100.0 480 1 US-08-440-210-14 Sequence 14, Appl
29 20 100.0 480 4 US-09-046-604-14 Sequence 14, Appl
30 20 100.0 621 1 US-07-748-292-7 Sequence 7, Appl
31 20 100.0 622 3 PCT-US92-06965A-4 Sequence 34, Appl
32 20 100.0 622 5 PCT-US92-06965A-4 Sequence 4, Appl
33 20 100.0 663 3 US-08-824-057-3 Sequence 3, Appl
34 20 100.0 663 4 US-09-415-582-3 Sequence 3, Appl
35 20 100.0 663 4 US-09-693-596-4 Sequence 35, Appl
36 20 100.0 738 3 US-08-867-611-35 Sequence 5, Appl
37 20 100.0 738 5 PCT-US92-06965A-5 Sequence 75, Appl
38 20 100.0 2436 3 US-08-444-818-75 Sequence 89, Appl
39 20 100.0 2772 3 US-08-444-818-89 Sequence 23, Appl
40 20 100.0 2894 2 US-08-466-975A-23 Sequence 23, Appl
41 20 100.0 2894 2 US-08-391-671A-23 Sequence 23, Appl
42 20 100.0 2894 3 US-08-467-902A-23 Sequence 23, Appl
43 20 100.0 2894 3 US-09-275-265-23 Sequence 23, Appl
44 20 100.0 2894 4 US-09-941-611-23 Sequence 3, Appl
45 20 100.0 2955 2 US-08-443-260-3 Sequence 3, Appl

ALIGNMENTS

RESULT 1

US-08-612-973-89
; Sequence 89, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-612-973-89

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTFKIRMYGVGEH 20

|||||

```
Db      1  PCTINTYIFKIRMYGVGVH 20

RESULT 2
US-08-927-597-89
; Sequence 89, Application US/08927597
; Patent No. 6245503
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-927-597-89

Query Match      100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  PCTINTYIFKIRMYGVGVH 20
Db      1  PCTINTYIFKIRMYGVGVH 20

RESULT 3
US-08-635-886C-45
; Sequence 45, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
```

```
Db      1  PCTINTYIFKIRMYGVGVH 20
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-45

Query Match      100.0%; Score 20; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  PCTINTYIFKIRMYGVGVH 20
Db      1  PCTINTYIFKIRMYGVGVH 20

RESULT 4
US-08-635-886C-250
; Sequence 250, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 250
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-250

Query Match      100.0%; Score 20; DB 4; Length 68;
Best Local Similarity 100.0%; Pred. No. 5.6e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  PCTINTYIFKIRMYGVGVH 20
Db      49  PCTINTYIFKIRMYGVGVH 68

RESULT 5
US-08-440-103-26
; Sequence 26, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,103
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-103-26

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTFKIRMYGVGVH 20
Db 66 PCTINYTFKIRMYGVGVH 85

RESULT 6
US-08-440-103-27
; Sequence 27, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,103
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-103-27

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTFKIRMYGVGVH 20
Db 66 PCTINYTFKIRMYGVGVH 85

RESULT 7
US-08-440-542-26
; Sequence 26, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,542
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-542-26

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTFKIRMYGVGVH 20
Db 66 PCTINYTFKIRMYGVGVH 85
```

```

; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-103-27

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTFKIRMYGVGVH 20
Db 66 PCTINYTFKIRMYGVGVH 85

RESULT 7
US-08-440-542-26
; Sequence 26, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,542
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-542-26

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTFKIRMYGVGVH 20
Db 66 PCTINYTFKIRMYGVGVH 85
```

```
RESULT 8
US-08-440-542-27
; Sequence 27, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,542
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-542-27

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINTYTFKIRMYGVGVH 20
Db 66 PCTINTYTFKIRMYGVGVH 85

RESULT 9
US-08-231-368-26
; Sequence 26, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-231-368-26

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINTYTFKIRMYGVGVH 20
Db 66 PCTINTYTFKIRMYGVGVH 85

RESULT 10
US-08-231-368-27
; Sequence 27, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
```

```
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-231-368-27

Query Match      100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINTYTFKIRMYGVGEH 20
Db 66 PCTINTYTFKIRMYGVGEH 85

RESULT 11
US-08-440-210-26
; Sequence 26, Application US/08440210
; Patent No. 576845
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,210
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 27:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-210-27

Query Match      100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINTYTFKIRMYGVGEH 20
Db 66 PCTINTYTFKIRMYGVGEH 85

RESULT 13
US-09-046-604-26
; Sequence 26, Application US/09046604
; Patent No. 6303292
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/046,604
;; FILING DATE: 13-SEP-1991
;; CLASSIFICATION: 13-SEP-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/231,368
;; FILING DATE: 13-SEP-1991
;; APPLICATION NUMBER: US 07/759,575
;; FILING DATE: 13-SEP-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: McClung, Barbara G.
;; REGISTRATION NUMBER: 33,113
;; REFERENCE/DOCKET NUMBER: 0205.001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (510) 601-2708
;; TELEFAX: (510) 655-3542
;; INFORMATION FOR SEQ ID NO: 26:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 88 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-09-046-604-26

Query Match 100.0%; Score 20; DB 4; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIKIRMYGVGVH 20
Db 66 PCTINYTIKIRMYGVGVH 85

RESULT 14
US-09-046-604-27
;; Sequence 27, Application US/09046604
;; Patent No. 6303292
;; GENERAL INFORMATION:
;; APPLICANT: Weiner, Amy J.
;; APPLICANT: Houghton, Michael
;; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
;; NUMBER OF SEQUENCES: 45
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Chiron Corporation
;; STREET: 4560 Horton Street
;; CITY: Emeryville
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94608
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/046,604
;; FILING DATE: 13-SEP-1991
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/231,368
;; FILING DATE: 13-SEP-1991
;; APPLICATION NUMBER: US 07/759,575
;; FILING DATE: 13-SEP-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: McClung, Barbara G.
;; REGISTRATION NUMBER: 33,113
;; REFERENCE/DOCKET NUMBER: 0205.001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (510) 601-2708
;; TELEFAX: (510) 655-3542
;; INFORMATION FOR SEQ ID NO: 27:
;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 88 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-09-046-604-27

Query Match 100.0%; Score 20; DB 4; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIKIRMYGVGVH 20
Db 66 PCTINYTIKIRMYGVGVH 85

RESULT 15
US-07-748-292-1
;; Sequence 1, Application US/07748292
;; Patent No. 5308750
;; GENERAL INFORMATION:
;; APPLICANT: MEHTA, SMRITI U.
;; APPLICANT: JOHNSON, JILL E.
;; APPLICANT: DAILEY, STEPHEN H.
;; APPLICANT: DESAI, SURESH M.
;; APPLICANT: DEVARÉ, SUSHIL G.
;; TITLE OF INVENTION: MONOCLONAL ANTIBODIES TO PUTATIVE HCV
;; NUMBER OF SEQUENCES: 10
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: ABBOTT LABORATORIES D-377/AP6D
;; STREET: ONE ABBOTT PARK ROAD
;; CITY: ABBOTT PARK
;; STATE: ILLINOIS
;; COUNTRY: USA
;; ZIP: 60064-3500
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/748,292
;; FILING DATE: 19910821
;; CLASSIFICATION: 436
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/456,162
;; FILING DATE: 22-DEC-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/610,180
;; FILING DATE: 07-NOV-1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: POREMSKI, PRISCILLA E.
;; REGISTRATION NUMBER: 33,207
;; REFERENCE/DOCKET NUMBER: 4767.US.P2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 708-937-6365
;; TELEFAX: 708-937-9556
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 121 amino acids
;; TYPE: AMINO ACID
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; US-07-748-292-1

Query Match 100.0%; Score 20; DB 1; Length 121;
Best Local Similarity 100.0%; Pred. No. 9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIKIRMYGVGVH 20
Db 66 PCTINYTIKIRMYGVGVH 85

Db 20 PCTINTFKIRMYGGVEH 39

Search completed: November 21, 2003, 21:15:18
Job time : 11.6 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time 19.5 Seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-89

Perfect score: 20

Sequence: 1 PCTINTYIKIRMYGVGEH 20

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	10	US-09-973-025-89
2	20	100.0	20	11	US-09-899-303-89
3	20	100.0	20	11	US-09-995-808-89
4	20	100.0	20	11	US-09-995-860-89
5	20	100.0	20	12	US-09-995-791-89
6	20	100.0	254	10	US-09-407-430-3
7	20	100.0	363	10	US-09-407-430-2
8	20	100.0	637	12	US-10-187-257-4
9	20	100.0	637	12	US-10-265-083-2
10	20	100.0	2894	10	US-09-941-611-23
11	20	100.0	2894	15	US-10-044-995-23
12	20	100.0	3011	9	US-09-916-359-2
13	20	100.0	3011	16	US-10-232-643-6
14	16	80.0	2940	12	US-10-226-629A-13
15	14	70.0	347	11	US-09-194-949-9
					Sequence 89, Appl
					Sequence 89, Appl
					Sequence 89, Appl
					Sequence 89, Appl
					Sequence 3, Appl
					Sequence 2, Appl
					Sequence 4, Appl
					Sequence 2, Appl
					Sequence 23, Appl
					Sequence 23, Appl
					Sequence 6, Appl
					Sequence 13, Appl
					Sequence 9, Appl

16 70.0 539 11 US-09-194-949-11 Sequence 11, Appl
17 10 50.0 176 10 US-09-921-397-81 Sequence 81, Appl
18 10 50.0 250 10 US-09-952-572-8 Sequence 8, Appl
19 10 50.0 255 10 US-09-894-018-89 Sequence 89, Appl
20 10 50.0 281 10 US-09-894-018-91 Sequence 91, Appl
21 10 50.0 350 10 US-09-929-955-4 Sequence 4, Appl
22 10 50.0 350 14 US-10-104-966-4 Sequence 4, Appl
23 10 50.0 363 12 US-10-128-587A-97 Sequence 97, Appl
24 10 50.0 363 15 US-10-128-590-97 Sequence 97, Appl
25 10 50.0 3011 9 US-09-742-659-4 Sequence 4, Appl
26 10 50.0 3011 10 US-09-238-076-20 Sequence 20, Appl
27 10 50.0 3011 10 US-09-932-572-9 Sequence 9, Appl
28 10 50.0 3011 10 US-09-929-955-1 Sequence 1, Appl
29 10 50.0 3011 10 US-09-747-419-20 Sequence 20, Appl
30 10 50.0 3011 11 US-09-891-894-3 Sequence 3, Appl
31 10 50.0 3011 11 US-09-995-937-20 Sequence 20, Appl
32 10 50.0 3011 11 US-09-917-563-20 Sequence 20, Appl
33 10 50.0 3011 12 US-10-184-150-3 Sequence 3, Appl
34 10 50.0 3011 14 US-10-104-966-1 Sequence 1, Appl
35 10 50.0 3011 15 US-10-259-275-20 Sequence 20, Appl
36 10 50.0 3012 11 US-09-238-076-2 Sequence 2, Appl
37 10 50.0 3012 11 US-09-995-937-2 Sequence 2, Appl
38 10 50.0 3012 11 US-09-917-563-2 Sequence 2, Appl
39 9 45.0 10 10 US-09-894-018-823 Sequence 223, Appl
40 9 45.0 28 16 US-10-318-200-17 Sequence 17, Appl
41 9 45.0 107 10 US-09-894-018-103 Sequence 103, Appl
42 9 45.0 130 10 US-09-894-018-99 Sequence 99, Appl
43 9 45.0 290 12 US-10-128-587A-3 Sequence 3, Appl
44 9 45.0 290 15 US-10-128-590-3 Sequence 3, Appl
45 9 45.0 301 12 US-10-128-587A-5 Sequence 5, Appl

ALIGNMENTS

RESULT 1

US-09-973-025-89
; Sequence 89, Application US/09973025
; Publication No. US20020182706A1

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

BOSMAN, FONS

DE MARTYNOFF, GUY

BUYSE, MARIE-ANG

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION NUMBER: US/09/973,025

FILING DATE: 10-Oct-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 89:
US-09-973-025-89

Query Match 100.0%; Score 20; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIKIRMYGVGVH 20
Db 1 PCTINYTIKIRMYGVGVH 20
|||||

RESULT 2
US-09-899-303-89
; Sequence 89, Application US/09899303
; Publication No. US20030036110A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 89:
US-09-899-303-89

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIKIRMYGVGVH 20
Db 1 PCTINYTIKIRMYGVGVH 20
|||||

Db 1 PCTINYTIKIRMYGVGVH 20

RESULT 3
US-09-995-808-89
; Sequence 89, Application US/09995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; therapeutic use.
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 89
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-89

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIKIRMYGVGVH 20
Db 1 PCTINYTIKIRMYGVGVH 20
|||||

RESULT 4
US-09-995-860-89
; Sequence 89, Application US/09995860
; Publication No. US20030118603A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; therapeutic use.
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 89
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-89

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIKIRMYGVGVH 20
Db 1 PCTINYTIKIRMYGVGVH 20
|||||

RESULT 5
US-09-995-791-89
; Sequence 89, Application US/09995791
; Publication No. US20030147918A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; therapeutic use.
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 89

```
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-89

Query Match      100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIKIRMYGGVEH 20
    |||||
Db 1 PCTINYTIKIRMYGGVEH 20
    |||||

RESULT 6
US-09-407-430-3
; Sequence 3, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-407-430-3

Query Match      100.0%; Score 20; DB 10; Length 254;
Best Local Similarity 100.0%; Pred. No. 1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIKIRMYGGVEH 20
    |||||
Db 214 PCTINYTIKIRMYGGVEH 233
    |||||

RESULT 7
US-09-407-430-2
; Sequence 2, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 363
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-09-407-430-2

Query Match      100.0%; Score 20; DB 10; Length 363;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIKIRMYGGVEH 20
    |||||
Db 236 PCTINYTIKIRMYGGVEH 255
    |||||

RESULT 8
US-10-187-257-4
; Sequence 4, Application US/10187257
; Publication No. US20030138458A1
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: O'HAGAN, Derek
; TITLE OF INVENTION: HCV E1E2 VACCINE COMPOSITIONS
; FILE REFERENCE: 2302-17206
; CURRENT APPLICATION NUMBER: US/10/187,257
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
US-10-187-257-4

Query Match      100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIKIRMYGGVEH 20
    |||||
Db 447 PCTINYTIKIRMYGGVEH 466
    |||||

RESULT 9
US-10-265-083-2
; Sequence 2, Application US/10265083
; Publication No. US20030170273A1
; GENERAL INFORMATION:
; APPLICANT: O'HAGAN, Derek
; APPLICANT: VALIANTE, Nicholas
; TITLE OF INVENTION: ADJUVANT COMPOSITIONS
; FILE REFERENCE: 2302-18357.30
; CURRENT APPLICATION NUMBER: US/10/265,083
; CURRENT FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7
US-10-265-083-2

Query Match      100.0%; Score 20; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIKIRMYGGVEH 20
    |||||
Db 447 PCTINYTIKIRMYGGVEH 466
    |||||

RESULT 10
US-09-941-611-23
; Sequence 23, Application US/09941611
; Patent No. US20020106640A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSWIN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; ANTIBODIES TO HEPATITIS C VIRUS
```

```
;
;
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/941,611
; FILING DATE: 30-Aug-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/391,671
; FILING DATE: 1995-02-21
; APPLICATION NUMBER: WO PCT/EP91/02409
; FILING DATE: 13-DEC-1991
; APPLICATION NUMBER: EP 90124241.2
; FILING DATE: 14-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 1487-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 7038164000
; TELEFAX: 7038164100
;
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
;
; US-09-941-611-23
;
; Query Match 100.0%; Score 20; DB 10; Length 2894;
; Best Local Similarity 100.0%; Pred. No. 8.5e-13;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 PCTINYTFKIRMYGVGVH 20
; Db 619 PCTINYTFKIRMYGVGVH 638
;
; RESULT 11
; US-10-044-995-23
; Sequence 23, Application US/10044995
; Publication No. US20030049685A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; POLLET, DIRK
; MAERTENS, GEERT
; VAN HEUVERSUN, HUGO
;
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; ANTIBODIES TO HEPATITIS C VIRUS
;
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
;
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/044,995
; FILING DATE: 15-Jan-2002
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/391,671
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 07/920,286
; FILING DATE: 14-OCT-1992
; APPLICATION NUMBER: WO PCT/EP91/02409
; FILING DATE: 13-DEC-1991
; APPLICATION NUMBER: EP 90124241.2
; FILING DATE: 14-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 1487-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 7038164000
; TELEFAX: 7038164100
;
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
;
; US-10-044-995-23
;
; Query Match 100.0%; Score 20; DB 15; Length 2894;
; Best Local Similarity 100.0%; Pred. No. 8.5e-13;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 PCTINYTFKIRMYGVGVH 20
; Db 619 PCTINYTFKIRMYGVGVH 638
;
; RESULT 12
; US-09-916-359-2
; Sequence 2, Application US/09916359
; Patent No. US20020034734A1
; GENERAL INFORMATION:
; APPLICANT: Veronique Barban
; TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
; TREATING C HEPATITIS
; FILE REFERENCE: PMCF97-03A
; CURRENT APPLICATION NUMBER: US/09/916,359
; CURRENT FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 09/388,874
; PRIOR FILING DATE: 1999-09-02
; PRIOR APPLICATION NUMBER: 97/02,887
; PRIOR FILING DATE: 1997-03-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Virus
;
; US-09-916-359-2
;
; Query Match 100.0%; Score 20; DB 9; Length 3011;
; Best Local Similarity 100.0%; Pred. No. 8.8e-13;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 PCTINYTFKIRMYGVGVH 20
; Db 619 PCTINYTFKIRMYGVGVH 638
```

Db 619 PCTINTIFKIRMYGVGEH 638

RESULT 13
US-10-232-643-6
: Sequence 6, Application US/10232643
: Publication No. US20030129586A1
: GENERAL INFORMATION:
: APPLICANT: CHOONG, MICHAEL
: CHOO, QUI-LIM
: HAN, JANG
: CHOE, JOONHO

TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
HELICASE ACTIVITY AND IMPROVED SOLUBILITY

NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHIRON CORPORATION
STREET: Intellectual Property - R440, P.O. Box 8097
CITY: Emeryville
STATE: California
COUNTRY: U.S.A.
ZIP: 94662-8097

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/232,643
FILING DATE: 30-Aug-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/483,799
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/529,169
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0100.005

TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 923-3274
TELEFAX: (510) 655-3542
TELEX: n/a

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3011 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

FEATURE:
NAME/KEY: Duplication
LOCATION: 9
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Lys or Arg"

FEATURE:
NAME/KEY: Duplication
LOCATION: 11
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Asn or Thr"

FEATURE:
NAME/KEY: Duplication
LOCATION: 176
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Ile or Thr"

FEATURE:
NAME/KEY: Duplication
LOCATION: 334
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Met or Val"

FEATURE:
NAME/KEY: Duplication
LOCATION: 2502
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Leu or Phe"

FEATURE:
NAME/KEY: Duplication
LOCATION: 603
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Leu or Ile"

FEATURE:
NAME/KEY: Duplication
LOCATION: 848
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Tyr or Asn"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1114
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Pro or Ser"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1117
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Ser or Thr"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1276
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Pro or Leu"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1454
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Cys or Tyr"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1471
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Thr or Ser"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1877
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Glu or Gly"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1948
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Leu or His"

FEATURE:
NAME/KEY: Duplication
LOCATION: 1949
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Ser or Cys"

FEATURE:
NAME/KEY: Duplication
LOCATION: 2021
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Gly or Val"

FEATURE:
NAME/KEY: Duplication
LOCATION: 2349
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Thr or Ser"

FEATURE:
NAME/KEY: Duplication
LOCATION: 2385
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Tyr or Phe"

FEATURE:
NAME/KEY: Duplication
LOCATION: 2386
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Ser or Ala"

FEATURE:
NAME/KEY: Duplication
LOCATION: 2502
OTHER INFORMATION: /note= "There exists a
heterogeneity at this position - Xaa = Leu or Phe"

```
;
;
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2690
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Arg or Gly"
;
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2921
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Arg or Gly"
;
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2996
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Leu or Pro"
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-232-643-6

Query Match      100.0%; Score 20; DB 16; Length 3011;
Best Local Similarity 100.0%; Pred. No. 8.8e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 PCTINTIFKIRMYVGVEH 20
Db      619 PCTINTIFKIRMYVGVEH 638
|||||
|||||

RESULT 14
US-10-226-629A-13
; Sequence 13, Application US/10226629A
; Publication No. US20030166504A1
; GENERAL INFORMATION:
; APPLICANT: Myriad Genetics, Inc.
; APPLICANT: Morham, Scott
; APPLICANT: Zavitz, Kenton
; APPLICANT: Hobden, Adrian
; TITLE OF INVENTION: Therapeutic Compositions and Methods for Treating Viral Infection
; FILE REFERENCE: 5006.01
; CURRENT APPLICATION NUMBER: US/10/226,629A
; CURRENT FILING DATE: 2002-08-22
; PRIOR APPLICATION NUMBER: US 60/314,182
; PRIOR FILING DATE: 2001-08-22
; NUMBER OF SEQ ID NOS: 736
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 2940
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-226-629A-13

Query Match      80.0%; Score 16; DB 12; Length 2940;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 NYTIFKIRMYVGVEH 20
Db      627 NYTIFKIRMYVGVEH 642
|||||
|||||

RESULT 15
US-09-194-949-9
; Sequence 9, Application US/09194949
; Publication No. US20030053987A1
; GENERAL INFORMATION:
; APPLICANT: Merck & Co., Inc.
; APPLICANT: Donnelly, John J.
; APPLICANT: Fu, Tong-Ming
; APPLICANT: Liu, Margaret A.
; APPLICANT: Shiver, John W.
; TITLE OF INVENTION: SYNTHETIC HEPATITIS C GENES
; FILE REFERENCE: 19732YP
; CURRENT APPLICATION NUMBER: US/09/194,949
; CURRENT FILING DATE: 2000-02-17

;
;
; PRIOR APPLICATION NUMBER: PCT/US97/09884
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/020,494
; PRIOR FILING DATE: 1996-06-11
; PRIOR APPLICATION NUMBER: 60/033,534
; PRIOR FILING DATE: 1996-12-20
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-09-194-949-9

Query Match      70.0%; Score 14; DB 11; Length 347;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 TIFKIRMYVGVEH 20
Db      243 TIFKIRMYVGVEH 256
|||||
|||||

Search completed: November 21, 2003, 22:19:41
Job time : 20.5 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 163.1 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-89

Perfect score: 20
Sequence: 1 PCTINTYIKIRMYGVGEH 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0
Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA Main: +
1: /cgn2_6/ptodata/1/paa/pctus COMB.pcp.*
2: /cgn2_6/ptodata/1/paa/us06 COMB.pcp.*
3: /cgn2_6/ptodata/1/paa/us07 COMB.pcp.*
4: /cgn2_6/ptodata/1/paa/us08 COMB.pcp.*
5: /cgn2_6/ptodata/1/paa/us081 COMB.pcp.*
6: /cgn2_6/ptodata/1/paa/us082 COMB.pcp.*
7: /cgn2_6/ptodata/1/paa/us083 COMB.pcp.*
8: /cgn2_6/ptodata/1/paa/us084 COMB.pcp.*
9: /cgn2_6/ptodata/1/paa/us085 COMB.pcp.*
10: /cgn2_6/ptodata/1/paa/us086 COMB.pcp.*
11: /cgn2_6/ptodata/1/paa/us087 COMB.pcp.*
12: /cgn2_6/ptodata/1/paa/us088 COMB.pcp.*
13: /cgn2_6/ptodata/1/paa/us089 COMB.pcp.*
14: /cgn2_6/ptodata/1/paa/us090 COMB.pcp.*
15: /cgn2_6/ptodata/1/paa/us091 COMB.pcp.*
16: /cgn2_6/ptodata/1/paa/us092 COMB.pcp.*
17: /cgn2_6/ptodata/1/paa/us093 COMB.pcp.*
18: /cgn2_6/ptodata/1/paa/us094 COMB.pcp.*
19: /cgn2_6/ptodata/1/paa/us095 COMB.pcp.*
20: /cgn2_6/ptodata/1/paa/us096 COMB.pcp.*
21: /cgn2_6/ptodata/1/paa/us097A COMB.pcp.*
22: /cgn2_6/ptodata/1/paa/us097B COMB.pcp.*
23: /cgn2_6/ptodata/1/paa/us098 COMB.pcp.*
24: /cgn2_6/ptodata/1/paa/us099A COMB.pcp.*
25: /cgn2_6/ptodata/1/paa/us099B COMB.pcp.*
26: /cgn2_6/ptodata/1/paa/us100 COMB.pcp.*
27: /cgn2_6/ptodata/1/paa/us101 COMB.pcp.*
28: /cgn2_6/ptodata/1/paa/us102 COMB.pcp.*
29: /cgn2_6/ptodata/1/paa/us103 COMB.pcp.*
30: /cgn2_6/ptodata/1/paa/us104 COMB.pcp.*
31: /cgn2_6/ptodata/1/paa/us106 COMB.pcp.*
32: /cgn2_6/ptodata/1/paa/us60 COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	20	100.0	20	13	US-08-974-685-45
					Sequence 45, Appl

2	20	100.0	20	13	US-08-974-690-45	Sequence 45, Appl
3	20	100.0	20	13	US-08-974-690A-45	Sequence 45, Appl
4	20	100.0	20	13	US-08-974-690B-45	Sequence 45, Appl
5	20	100.0	20	13	US-08-974-690C-45	Sequence 45, Appl
6	20	100.0	20	23	US-09-899-303-89	Sequence 89, Appl
7	20	100.0	20	23	US-09-899-303A-89	Sequence 89, Appl
8	20	100.0	20	25	US-09-973-025-89	Sequence 89, Appl
9	20	100.0	20	25	US-09-995-791-89	Sequence 89, Appl
10	20	100.0	20	25	US-09-995-808-89	Sequence 89, Appl
11	20	100.0	20	25	US-09-995-860-89	Sequence 89, Appl
12	20	100.0	20	26	US-10-020-510-89	Sequence 89, Appl
13	20	100.0	20	29	US-10-321-798-89	Sequence 89, Appl
14	20	100.0	68	13	US-08-974-690C-250	Sequence 250, Appl
15	20	100.0	88	8	US-08-471-498-26	Sequence 26, Appl
16	20	100.0	88	8	US-08-471-498-27	Sequence 27, Appl
17	20	100.0	179	8	US-08-403-590B-77	Sequence 77, Appl
18	20	100.0	179	8	US-08-444-112-77	Sequence 77, Appl
19	20	100.0	254	1	PCT-US00-26395-3	Sequence 3, Appl
20	20	100.0	254	18	US-09-407-430-3	Sequence 3, Appl
21	20	100.0	333	30	US-10-445-724-3	Sequence 3, Appl
22	20	100.0	333	32	US-60-409-909-4	Sequence 4, Appl
23	20	100.0	363	1	PCT-US00-26395-2	Sequence 2, Appl
24	20	100.0	363	18	US-09-407-430-2	Sequence 2, Appl
25	20	100.0	414	3	US-07-748-561-2	Sequence 2, Appl
26	20	100.0	454	8	US-08-403-590B-73	Sequence 73, Appl
27	20	100.0	454	8	US-08-444-112-73	Sequence 73, Appl
28	20	100.0	480	8	US-08-471-498-14	Sequence 14, Appl
29	20	100.0	531	21	US-09-721-480-7	Sequence 7, Appl
30	20	100.0	522	3	US-07-748-561-4	Sequence 4, Appl
31	20	100.0	622	3	US-07-989-843-34	Sequence 34, Appl
32	20	100.0	622	8	US-08-463-849-34	Sequence 34, Appl
33	20	100.0	622	8	US-08-463-884-34	Sequence 34, Appl
34	20	100.0	622	20	US-09-690-359-34	Sequence 34, Appl
35	20	100.0	637	1	PCT-US02-20676-4	Sequence 4, Appl
36	20	100.0	637	1	PCT-US02-30423-2	Sequence 2, Appl
37	20	100.0	637	1	PCT-US02-31486-2	Sequence 2, Appl
38	20	100.0	637	27	US-10-187-257-4	Sequence 4, Appl
39	20	100.0	637	28	US-10-265-083-2	Sequence 2, Appl
40	20	100.0	663	6	US-08-282-959-3	Sequence 3, Appl
41	20	100.0	663	14	US-09-073-406-4	Sequence 4, Appl
42	20	100.0	663	29	US-10-371-040-4	Sequence 4, Appl
43	20	100.0	732	3	US-07-748-561-5	Sequence 5, Appl
44	20	100.0	738	3	US-07-989-843-35	Sequence 35, Appl
45	20	100.0	738	8	US-08-463-849-35	Sequence 35, Appl

ALIGNMENTS

RESULT 1

US-08-974-685-45

; Sequence 45, Application US/08974685

; GENERAL INFORMATION:

; APPLICANT: LEROUX-ROELS, GEERT

; DELEYS, ROBERT

; MAERTENS, GEERT

; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF

; HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

; THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

; NUMBER OF SEQUENCES: 181

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: NIXON & VANDERHVE P.C.

; STREET: 1100 NORTH GLEBE ROAD

; CITY: ARLINGTON

; STATE: VIRGINIA

; COUNTRY: U.S.A.

; ZIP: 22201-4714

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,685
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-08-974-685-45

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PCTINTIFKIRMYGVGVH 20
Db 1 PCTINTIFKIRMYGVGVH 20

RESULT 2
US-08-974-690-45
Sequence 45, Application US/08974690
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
APPLICANT: MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 166
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/635,886
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4100
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

US-08-974-690-45

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PCTINTIFKIRMYGVGVH 20
Db 1 PCTINTIFKIRMYGVGVH 20

RESULT 3
US-08-974-690A-45
Sequence 45, Application US/08974690A
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
APPLICANT: MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND
THERAPEUTIC COMPOSITIONS CONTAINING THE SAME
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690A
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-08-974-690A-45

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PCTINTIFKIRMYGVGVH 20
Db 1 PCTINTIFKIRMYGVGVH 20

RESULT 4
US-08-974-690B-45
Sequence 45, Application US/08974690B
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, GEERT
APPLICANT: DELEYS, ROBERT
APPLICANT: MAERTENS, GEERT
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF
HEPATITIS C VIRUS STRUCTURAL PROTEINS, PROPHYLACTIC AND

THIS PAGE BLANK (USPTO)

Db 21 GERCLEDRDRSEL 34
|||||

RESULT 13

Q9E3F4 PRELIMINARY; PRT; 74 AA.
AC Q9E3F4; STRAIN=2-16;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2 glycoprotein-PKR-eIF-2 (genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; sRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2-16;
RX MEDLINE=20375028; PubMed=10915068;
RA Polyak S.J., Nausbaum J.B., Larson A.M., Cotler S., Carithers Jr R.L.,
RA Gretch D.R.;
RT "The protein kinase-interacting domain in the hepatitis C virus
RT envelope glycoprotein-2 gene is highly conserved in genotype 1-
RT infected patients treated with interferon.";
RL J. Infect. Dis. 182:397-404(2000).
DR EMBL; AF264984; AAG17323.1; -;
DR InterPro; IPR002531; HCV_NSI.1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 74
SQ SEQUENCE 74 AA; 8609 MW; 655F1D877C1F8049 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 74;
Best Local Similarity 100.0%; Pred. No. 6.4e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
|||||

Db 21 GERCLEDRDRSEL 34

RESULT 14

Q9E3H2 PRELIMINARY; PRT; 74 AA.
AC Q9E3H2; STRAIN=11w6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2 glycoprotein-PKR-eIF-2 (genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; sRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=11w6;
RX MEDLINE=20375028; PubMed=10915068;
RA Polyak S.J., Nausbaum J.B., Larson A.M., Cotler S., Carithers Jr R.L.,
RA Gretch D.R.;
RT "The protein kinase-interacting domain in the hepatitis C virus
RT envelope glycoprotein-2 gene is highly conserved in genotype 1-
RT infected patients treated with interferon.";
RL J. Infect. Dis. 182:397-404(2000).
DR EMBL; AF264966; AAG17305.1; -;
DR InterPro; IPR002531; HCV_NSI.1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.

FT NON_TER 1
FT NON_TER 74
SQ SEQUENCE 74 AA; 8469 MW; 01E4DD9B0DFD1333 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 74;
Best Local Similarity 100.0%; Pred. No. 6.4e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
|||||

Db 21 GERCLEDRDRSEL 34

RESULT 15

Q9E3G0 PRELIMINARY; PRT; 74 AA.
AC Q9E3G0; STRAIN=2-8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2 glycoprotein-PKR-eIF-2 (genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; sRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2-8;
RX MEDLINE=20375028; PubMed=10915068;
RA Polyak S.J., Nausbaum J.B., Larson A.M., Cotler S., Carithers Jr R.L.,
RA Gretch D.R.;
RT "The protein kinase-interacting domain in the hepatitis C virus
RT envelope glycoprotein-2 gene is highly conserved in genotype 1-
RT infected patients treated with interferon.";
RL J. Infect. Dis. 182:397-404(2000).
DR EMBL; AF264978; AAG17317.1; -;
DR InterPro; IPR002531; HCV_NSI.1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 74
SQ SEQUENCE 74 AA; 8438 MW; 7C651D9B130CFC26 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 74;
Best Local Similarity 100.0%; Pred. No. 6.4e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
|||||

Db 21 GERCLEDRDRSEL 34

Search completed: November 21, 2003, 21:08:22
Job time : 24.45 secs

```

OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SR20;
RX MEDLINE=20287468; PubMed=10827164;
RA Sarrazin C., Kornetzky I., Ruster B., Lee J.H., Kronenberger B.,
RA Bruch K., Roth W.K., Zeuzem S.;
RT "Mutations within E2 and NS5A protein in patients infected with
RT Hepatitis C virus type 3a and correlation with treatment response.";
RL Hepatology 31:1360-1370(2000).
DR EMBL; AJ289090; CAB91861.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; I.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 67 AA; 7522 MW; B54C6455B9B3A996 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 67;
Best Local Similarity 100.0%; Pred. No. 5.9e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
Db 18 GERCLEDRDRSEL 31

RESULT 10
Q91EJ4 PRELIMINARY; PRT; 67 AA.
AC Q91EJ4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Envelope protein 2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ETR8;
RX MEDLINE=20287468; PubMed=10827164;
RA Sarrazin C., Kornetzky I., Ruster B., Lee J.H., Kronenberger B.,
RA Bruch K., Roth W.K., Zeuzem S.;
RT "Mutations within E2 and NS5A protein in patients infected with
RT Hepatitis C virus type 3a and correlation with treatment response.";
RL Hepatology 31:1360-1370(2000).
DR EMBL; AJ289093; CAB91864.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; I.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 67 AA; 7578 MW; ABAB9D55AF5A996 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 67;
Best Local Similarity 100.0%; Pred. No. 5.9e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
Db 18 GERCLEDRDRSEL 31

RESULT 11
Q91EJ9 PRELIMINARY; PRT; 67 AA.
ID Q91EJ9
AC Q91EJ9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Envelope protein 2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ETR8;
RX MEDLINE=20287468; PubMed=10827164;
RA Sarrazin C., Kornetzky I., Ruster B., Lee J.H., Kronenberger B.,
RA Bruch K., Roth W.K., Zeuzem S.;
RT "Mutations within E2 and NS5A protein in patients infected with
RT Hepatitis C virus type 3a and correlation with treatment response.";
RL Hepatology 31:1360-1370(2000).
DR EMBL; AJ289093; CAB91864.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; I.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 67 AA; 7578 MW; ABAB9D55AF5A996 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 67;
Best Local Similarity 100.0%; Pred. No. 5.9e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
Db 18 GERCLEDRDRSEL 31

RESULT 12
Q9E3I1 PRELIMINARY; PRT; 73 AA.
ID Q9E3I1
AC Q9E3I1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2 glycoprotein-PKR-eIF-2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5WK6;
RX MEDLINE=20375028; PubMed=10915068;
RA Polyak S.J., Nausbaum J.B., Larson A.M., Cotler S., Carithers Jr R.L.,
RA Gretch D.R.;
RT "The protein kinase-interacting domain in the hepatitis C virus
RT envelope glycoprotein-2 gene is highly conserved in genotype 1-
RT infected patients treated with interferon.";
RL J. Infect. Dis. 182:397-404(2000).
DR EMBL; AF264957; AAG17296.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; I.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 73 AA; 8340 MW; 251D9B130CFC260C CRC64;

Query Match 70.0%; Score 14; DB 12; Length 73;
Best Local Similarity 100.0%; Pred. No. 6.4e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20

```

KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 67
 SQ SEQUENCE 67 AA; B5517955A4AEA996 CRC64;
 Query Match 70.0%; Score 14; DB 12; Length 67;
 Best Local Similarity 100.0%; Pred. No. 5.9e-08;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 GERCLEDRDRSEL 20
 Db 18 GERCLEDRDRSEL 31
 RESULT 6
 Q9IEJ9 PRELIMINARY; PRT; 67 AA.
 ID Q9IEJ9
 AC Q9IEJ9
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Envelope protein 2 (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=111103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NR13;
 RX MEDLINE=20287468; PubMed=10827164;
 RA Sarrazin C., Kornatzky I., Ruster B., Lee J.H., Kronenberger B.,
 RA Bruch K., Roth W.K., Zeuzem S.;
 RT "Mutations within E2 and NS5A protein in patients infected with
 RT hepatitis C virus type 3a and correlation with treatment response.";
 RL Hepatology 31:1360-1370(2000).
 DR EMBL; AJ289098; CAB91869.1; -.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1.
 DR ProbDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 67
 SQ SEQUENCE 67 AA; B5517955A4AEA996 CRC64;
 Query Match 70.0%; Score 14; DB 12; Length 67;
 Best Local Similarity 100.0%; Pred. No. 5.9e-08;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 GERCLEDRDRSEL 20
 Db 18 GERCLEDRDRSEL 31
 RESULT 7
 Q9IEJ0 PRELIMINARY; PRT; 67 AA.
 ID Q9IEJ0
 AC Q9IEJ0
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Envelope protein 2 (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=111103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NR12;
 RX MEDLINE=20287468; PubMed=10827164;
 RA Sarrazin C., Kornatzky I., Ruster B., Lee J.H., Kronenberger B.,
 RA Bruch K., Roth W.K., Zeuzem S.;

RT "Mutations within E2 and NS5A protein in patients infected with
 RT hepatitis C virus type 3a and correlation with treatment response.";
 RL Hepatology 31:1360-1370(2000).
 DR EMBL; AJ289097; CAB91868.1; -.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProbDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 67
 SQ SEQUENCE 67 AA; B5517955A4AEA996 CRC64;
 Query Match 70.0%; Score 14; DB 12; Length 67;
 Best Local Similarity 100.0%; Pred. No. 5.9e-08;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 GERCLEDRDRSEL 20
 Db 18 GERCLEDRDRSEL 31
 RESULT 8
 Q9IEJ3 PRELIMINARY; PRT; 67 AA.
 ID Q9IEJ3
 AC Q9IEJ3
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Envelope protein 2 (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=111103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ETR9;
 RX MEDLINE=20287468; PubMed=10827164;
 RA Sarrazin C., Kornatzky I., Ruster B., Lee J.H., Kronenberger B.,
 RA Bruch K., Roth W.K., Zeuzem S.;
 RT "Mutations within E2 and NS5A protein in patients infected with
 RT hepatitis C virus type 3a and correlation with treatment response.";
 RL Hepatology 31:1360-1370(2000).
 DR EMBL; AJ289094; CAB91865.1; -.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProbDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 67
 SQ SEQUENCE 67 AA; B5517955A4AEA996 CRC64;
 Query Match 70.0%; Score 14; DB 12; Length 67;
 Best Local Similarity 100.0%; Pred. No. 5.9e-08;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 GERCLEDRDRSEL 20
 Db 18 GERCLEDRDRSEL 31
 RESULT 9
 Q9IEJ7 PRELIMINARY; PRT; 67 AA.
 ID Q9IEJ7
 AC Q9IEJ7
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Envelope protein 2 (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.

Best Local Similarity 100.0%; Pred. No. 5.9e-08; Mismatches 0; Indels 0; Gaps 0;
Matches 14; Conservative 0;

QY 7 GERCDEDRDRSEL 20
| | | | | | | | | | | | | | | | | |

Db 18 GERCDEDRDRSEL 31
| | | | | | | | | | | | | | | | | |

RESULT 4

Q9IEJ5 PRELIMINARY; PRT; 67 AA.
ID Q9IEJ5
AC Q9IEJ5
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Envelope protein 2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ETR7;
RX MEDLINE=20287468; PubMed=10827164;
RA Sarrazin C., Kornatzky I., Ruster B., Lee J.H., Kronenberger B.,
Bruch K., Roth W.K., Zeuzem S.;
RT "Mutations within E2 and NS5A protein in patients infected with
Hepatitis C virus type 3a and correlation with treatment response.";
RL Hepatology 31:1360-1370(2000).
DR EMBL; AJ289092; CAB91863.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1 1
FT NON_TER 67 67
SQ SEQUENCE 67 AA; 7536 MW; ABB08B5B9B3A996 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 67;
Best Local Similarity 100.0%; Pred. No. 5.9e-08;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCDEDRDRSEL 20
| | | | | | | | | | | | | | | | | |

Db 18 GERCDEDRDRSEL 31
| | | | | | | | | | | | | | | | | |

RESULT 5

Q9IEJ2 PRELIMINARY; PRT; 67 AA.
ID Q9IEJ2
AC Q9IEJ2
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Envelope protein 2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ETR10;
RX MEDLINE=20287468; PubMed=10827164;
RA Sarrazin C., Kornatzky I., Ruster B., Lee J.H., Kronenberger B.,
Bruch K., Roth W.K., Zeuzem S.;
RT "Mutations within E2 and NS5A protein in patients infected with
Hepatitis C virus type 3a and correlation with treatment response.";
RL Hepatology 31:1360-1370(2000).
DR EMBL; AJ289095; CAB91866.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.

RESULT 2

Q9IEJ8 PRELIMINARY; PRT; 67 AA.
ID Q9IEJ8
AC Q9IEJ8
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Envelope protein 2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SR19;
RX MEDLINE=20287468; PubMed=10827164;
RA Sarrazin C., Kornatzky I., Ruster B., Lee J.H., Kronenberger B.,
Bruch K., Roth W.K., Zeuzem S.;
RT "Mutations within E2 and NS5A protein in patients infected with
Hepatitis C virus type 3a and correlation with treatment response.";
RL Hepatology 31:1360-1370(2000).
DR EMBL; AJ289089; CAB91860.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1 1
FT NON_TER 67 67
SQ SEQUENCE 67 AA; 7536 MW; ABB08B5B9B3A996 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 67;
Best Local Similarity 100.0%; Pred. No. 5.9e-08;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCDEDRDRSEL 20
| | | | | | | | | | | | | | | | | |

Db 18 GERCDEDRDRSEL 31
| | | | | | | | | | | | | | | | | |

RESULT 3

Q9IEI7 PRELIMINARY; PRT; 67 AA.
ID Q9IEI7
AC Q9IEI7
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Envelope protein 2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR15;
RX MEDLINE=20287468; PubMed=10827164;
RA Sarrazin C., Kornatzky I., Ruster B., Lee J.H., Kronenberger B.,
Bruch K., Roth W.K., Zeuzem S.;
RT "Mutations within E2 and NS5A protein in patients infected with
Hepatitis C virus type 3a and correlation with treatment response.";
RL Hepatology 31:1360-1370(2000).
DR EMBL; AJ289100; CAB91871.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1 1
FT NON_TER 67 67
SQ SEQUENCE 67 AA; 7549 MW; B5517955A4EA996 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 67;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:45:57 ; Search time 24.4 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-91

Perfect score: 20
Sequence: 1 ACNWTFGERCDLEDRSEL.20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	70.0	67	12 Q9IEK0	Q9iek0 hepatitis c
2	14	70.0	67	12 Q9IEJ8	Q9iej8 hepatitis c
3	14	70.0	67	12 Q9IEI7	Q9iei7 hepatitis c
4	14	70.0	67	12 Q9IEJ5	Q9iej5 hepatitis c
5	14	70.0	67	12 Q9IEJ2	Q9iej2 hepatitis c
6	14	70.0	67	12 Q9IEI9	Q9iei9 hepatitis c
7	14	70.0	67	12 Q9IEJ0	Q9iej0 hepatitis c
8	14	70.0	67	12 Q9IEJ3	Q9iej3 hepatitis c
9	14	70.0	67	12 Q9IEJ7	Q9iej7 hepatitis c
10	14	70.0	67	12 Q9IEJ4	Q9iej4 hepatitis c
11	14	70.0	67	12 Q9IEJ9	Q9iej9 hepatitis c
12	14	70.0	73	12 Q9E3I1	Q9e3i1 hepatitis c
13	14	70.0	74	12 Q9E3F4	Q9e3f4 hepatitis c
14	14	70.0	74	12 Q9E3H2	Q9e3h2 hepatitis c
15	14	70.0	74	12 Q9E3G0	Q9e3g0 hepatitis c
16	14	70.0	74	12 Q9E3H7	Q9e3h7 hepatitis c

17	14	70.0	74	12 Q9E3F9	Q9e3f9 hepatitis c
18	14	70.0	74	12 Q9E3G5	Q9e3g5 hepatitis c
19	14	70.0	74	12 Q9E3H8	Q9e3h8 hepatitis c
20	14	70.0	74	12 Q9E3G3	Q9e3g3 hepatitis c
21	14	70.0	74	12 Q9E3G6	Q9e3g6 hepatitis c
22	14	70.0	74	12 Q9E3F5	Q9e3f5 hepatitis c
23	14	70.0	74	12 Q9E3F7	Q9e3f7 hepatitis c
24	14	70.0	74	12 Q9E3G1	Q9e3g1 hepatitis c
25	14	70.0	74	12 Q9E3H1	Q9e3h1 hepatitis c
26	14	70.0	74	12 Q9E3G8	Q9e3g8 hepatitis c
27	14	70.0	74	12 Q9E3H9	Q9e3h9 hepatitis c
28	14	70.0	74	12 Q9E3G7	Q9e3g7 hepatitis c
29	14	70.0	74	12 Q9E3F8	Q9e3f8 hepatitis c
30	14	70.0	74	12 Q9E3F6	Q9e3f6 hepatitis c
31	14	70.0	74	12 Q9E3I0	Q9e3i0 hepatitis c
32	14	70.0	74	12 Q9E3H0	Q9e3h0 hepatitis c
33	14	70.0	74	12 Q9E3G4	Q9e3g4 hepatitis c
34	14	70.0	74	12 Q9E3H5	Q9e3h5 hepatitis c
35	14	70.0	74	12 Q9E3G2	Q9e3g2 hepatitis c
36	14	70.0	74	12 Q9E3H4	Q9e3h4 hepatitis c
37	14	70.0	85	12 Q9IJM2	Q9ijm2 hepatitis c
38	14	70.0	87	12 Q9IJP3	Q9ijp3 hepatitis c
39	14	70.0	91	12 Q9IJP6	Q9ijp6 hepatitis c
40	14	70.0	92	12 Q9IJM8	Q9ijm8 hepatitis c
41	14	70.0	94	12 Q9IJP9	Q9ijp9 hepatitis c
42	14	70.0	96	12 Q9IJM0	Q9ijm0 hepatitis c
43	14	70.0	97	12 Q9IJK4	Q9ijk4 hepatitis c
44	14	70.0	97	12 Q9IJJ3	Q9ijj3 hepatitis c
45	14	70.0	97	12 Q9IJN3	Q9ijn3 hepatitis c

ALIGNMENTS

RESULT 1

Q9IEK0 PRELIMINARY; PRT; 67 AA.
ID Q9IEK0;
AC Q9IEK0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Envelope protein 2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RC STRAIN=SR17;
RX MEDLINE=20287468; PubMed=10827164;
RA Sarrazin C., Kornetzy I., Ruster B., Lee J.H., Kronenberger B.,
RA Bruch K., Roth W.K., Zeuzem S.;
RT "Mutations within E2 and NS5A protein in patients infected with
RT hepatitis C virus type 3a and correlation with treatment response.";
RL Hepatology 31:1360-1370(2000).
DR EMBL; AJ289087; CAB91858.1; -
DR InterPro: IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 67
SQ SEQUENCE 67 AA; 7563 MW; ABA09655A4AEA996 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 67;
Best Local Similarity 100.0%; Pred. No. 5.9e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GERCLEDRSEL 20

Db 18 GERCLEDRSEL 31

THIS PAGE BLANK (11SPT0)

```
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AL022021; CAAL7707.1; -.
DR EMBL; AE007042; AAK46104.1; ALT_INIT.
DR PIR; C70929; C70929.
DR HSSP; Q55080; I107.
DR TIGR; MT1834; -.
DR Tuberculist; RV1785c; -.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW Complete proteome.
FT METAL 342 342 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 393 AA; 43541 MW; 8BCD1D50B471FAB8 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 393;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LEDRDR 17
DB 154 LEDRDR 159

RESULT 15
POL_AVIRE STANDARD; PRT; 473 AA.
AC F03360;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE POL polyprotein [Contains: Reverse transcriptase (EC 2.7.7.49);
DE Endonuclease] (Fragment).
GN POL.
OS Avian reticuloendotheliosis virus.
OC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=11636;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=A;
RX MEDLINE=85009850; PubMed=6090694;
RA Wilhelmsen K.C., Eggleston K., Temin H.M.;
RT "Nucleic acid sequences of the oncogene v-rel in
RT reticuloendotheliosis virus strain T and its cellular homolog, the
RT proto-oncogene c-rel.";
RL J. Virol. 52:172-182(1984).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA}(N).
CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- MISCELLANEOUS: STRAIN A IS A HELPER VIRUS OF THE STRAIN T.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; K02537; AAA99197.1; -.
DR EMBL; X01455; CAA25685.1; -.
DR PIR; A03959; A03959.
DR MEROPS; A02.008; -.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR Pfam; PF00075; rnaseh; 1.
DR Pfam; PF00665; rve; 1.
KW Transferase; RNA-directed DNA polymerase; Hydrolase; Nuclease;
KW Endonuclease; Polyprotein.
FT NON_TER 1 1
SQ SEQUENCE 473 AA; 52212 MW; 89FG2FA489EF51DD CRC64;
```

```
Query Match 30.0%; Score 6; DB 1; Length 473;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LEDRDR 17
DB 303 LEDRDR 308

Search completed: November 21, 2003, 21:00:03
Job time : 4.85 secs
```



```

CC -!- SIMILARITY: BELONGS TO THE AHPC/TSA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L14640; AAA16374.1; -
CC EMBL; Z46659; CAA86627.1; -
CC PIR; A47362; A47362.
CC SWISS-2DPAGE; P34760; YEAST.
CC SGD; S0004490; TSA1.
CC GO; GO:0005737; C:cytoplasm; IDA.
CC GO; GO:0008379; F:thioredoxin peroxidase activity; IDA.
CC GO; GO:0030503; P:regulation of redox homeostasis; IDA.
CC GO; GO:0006979; P:response to oxidative stress; IDA.
CC InterPro; IPR000866; Ahpc-TSA.
CC DR Pfam; PF00578; Ahpc-TSA; 1.
CC Antioxidant.
CC INIT MET 0 0 BY SIMILARITY.
CC ACT_SITE 47 47 BY SIMILARITY.
CC ACT_SITE 170 170 BY SIMILARITY.
CC ACT_SITE 170 170 BY SIMILARITY.
CC SQ SEQUENCE 195 AA; 21458 MW; 7DC61FE4BF3B564A CRC64;
Query Match 30.0%; Score 6; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 CNWTPG 7
DB 170 CNWTPG 175
RESULT 13
TSA2 YEAST
ID TSA2 YEAST STANDARD; PRT; 195 AA.
AC Q04120;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable thiol-specific antioxidant protein 2.
GN YDR453C OR D9461.38.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
ON NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E.,
RA Berno A., Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,
RA Hunnicke-Smith S., Hyman R., Komp C., Lashkari D., Lew H., Lin D.,
RA Moseedale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X.,
RA Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N.,
RA Winant A., Yelton M., Botstein D., Davis R.W.,
RA Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: HOMODIMER; DISULFIDE-LINKED, UPON OXIDATION (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- PTM: THE CYS-47-SH GROUP IS THE PRIMARY SITE OF OXIDATION BY
CC H2O(2), AND THE OXIDIZED CYS-47 (PROBABLY CYS-SOH) RAPIDLY
CC REACTS WITH CYS-170-SH OF THE OTHER SUBUNIT TO FORM AN
CC INTERMOLECULAR DISULFIDE. THIS DISULFIDE IS SUBSEQUENTLY REDUCED
CC BY THIOREDIXIN.
CC -!- SIMILARITY: BELONGS TO THE AHPC/TSA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U33007; AAB64486.1; -
CC PIR; S69732; S69732.
CC SGD; S0002561; YDR453C.
CC GO; GO:0005634; C:nucleus; IDA.
CC GO; GO:0008379; F:thioredoxin peroxidase activity; IDA.
CC GO; GO:0030503; P:regulation of redox homeostasis; IDA.
CC InterPro; IPR000866; Ahpc-TSA.
CC DR Pfam; PF00578; Ahpc-TSA; 1.
CC Antioxidant.
CC INIT MET 0 0 BY SIMILARITY.
CC ACT_SITE 47 47 BY SIMILARITY.
CC ACT_SITE 170 170 BY SIMILARITY.
CC ACT_SITE 170 170 BY SIMILARITY.
CC SQ SEQUENCE 195 AA; 21483 MW; A4C508C98F484145 CRC64;
Query Match 30.0%; Score 6; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 CNWTPG 7
DB 170 CNWTPG 175
RESULT 14
C143 MYCTU
ID C143 MYCTU STANDARD; PRT; 393 AA.
AC Q53936;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative cytochrome P450 143 (EC 1.14.-.-).
GN CYP143 OR RV1785C OR MT1834 OR MTV049.07C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
ON NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=H37RV;
RX MEDLINE=98255987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaita F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RA complete genome sequence."
RA Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann J.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RA laboratory strains."
RA Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

DE subunit).
GN CFBF OR PEBP2B OR PEBP2B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 12-28 AND 95-98.
RX MEDLINE=93242761; PubMed=8386878;
RA Ogawa E., Inuzuka M., Maruyama M., Satake M., Naito-Fujimoto M.,
RA Ito Y., Shigesada K.;
RT "Molecular Cloning and Characterization of PEBP2 beta, the
RT heterodimeric partner of a novel Drosophila runt-related DNA binding
RT protein PEBP2 alpha.";
RL Virology 194:314-331(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Thymus;
RX MEDLINE=93268282; PubMed=8497254;
RA Wang S., Wang Q., Crute B.E., Melnikova I.N., Keller S.R.,
RA Speck N.A.;
RT "Cloning and characterization of subunits of the T-cell receptor and
RT murine leukemia virus enhancer core-binding factor.";
RL Mol. Cell. Biol. 13:3324-3339(1993).
CC -!- FUNCTION: CBF BINDS TO THE CORE SITE, 5'-PYGPGGT-3', OF A NUMBER
CC OF ENHANCERS AND PROMOTERS, INCLUDING MURINE LEUKEMIA VIRUS,
CC POLYOMAVIRUS ENHANCER, T-CELL RECEPTOR ENHANCERS, LCK, IL-3 AND
CC GM-CSF PROMOTERS. THE BETA SUBUNIT ACTS TO ENHANCE THE DNA BINDING
CC OF THE ALPHA SUBUNIT.
CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1;
CC IsoId=Q08024-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q08024-2; Sequence=VSP_004360;
CC Note=Major isoform;
CC Name=3;
CC IsoId=Q08024-3; Sequence=VSP_004359;
CC Notes=Does not dimerize with the alpha subunit;
CC Name=4;
CC IsoId=Q08024-4; Sequence=VSP_004358;
CC Notes=Does not dimerize with the alpha subunit;
CC -!- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES TESTED. HIGHEST LEVEL
CC IN THYMUS, BUT ALSO ABUNDANTLY EXPRESSED IN MUSCLE, LUNG AND
CC BRAIN.
CC -!- SIMILARITY: BELONGS TO THE CBF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D14572; BAA03426.1; -
DR EMBL; D14571; BAA03425.1; -
DR EMBL; D14570; BAA03424.1; -
DR EMBL; D14569; BAA03423.1; -
DR EMBL; L03305; -; NOT ANNOTATED CDS.
DR EMBL; L03306; -; NOT ANNOTATED CDS.
DR EMBL; L03279; -; NOT_ANNOTATED_CDS.
DR PIR; A46107; A46107.
DR FIR; B48124; B48124.
DR PDB; 1JLF; 26-SEP-01.
DR PDB; 2JHB; 05-JUL-99.
DR TRANSFAC; T01064; -
DR TRANSFAC; T01065; -
DR TRANSFAC; T02254; -
DR TRANSFAC; T02255; -
MGD; MGI:99851; Cbfb.

DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0005515; P:protein binding activity; IPI.
DR GO; GO:0030098; P:lymphocytic blood cell differentiation; IMP.
DR GO; GO:0030099; P:myeloid blood cell differentiation; IMP.
DR GO; GO:0001503; P:osification; IMP.
DR GO; GO:0045944; P:positive regulation of transcription from P...; IDA.
DR InterPro; IPR003417; CBF_beta.
DR Pfam; PF02312; CBF_beta; 1.
KW Nuclear protein; Alternative splicing; Phosphorylation; 3D-structure.
FT MOD_RES 10 10 PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT MOD_RES 159 159 PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT VARSPLIC 56 94 Missing (in isoform 4).
FT VARSPLIC 134 165 Missing (in isoform 3).
FT VARSPLIC 166 187 Missing (in isoform 2).
FT VARSPLIC 171 171 AARQQDPSPGNSLGGDDLKLR -> VRVSQLLAVTGKKA
FT CONFLICT 171 171 /FTID=VSP_004360.
FT SEQUENCE 187 AA; 22030 MW; 0B2E6101A35D0FD8 CRC64;
Query Match 30.0%; Score 6; DB 1; Length 187;
Best Local Similarity 100.0%; Pred.No. 8.2;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 13 EDRDRS 18
DB 154 EDRDRS 159
|||||
RESULT 12
TSAL YEAST STANDARD; PRT; 195 AA.
ID TSAL YEAST
AC P34760;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Thiol-specific antioxidant protein (PRP).
GE TSAL OR TSA OR YML028W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=JD7-7C;
RX MEDLINE=93346439; PubMed=8344960;
RA Chae H.Z., Kim I.-H., Kim K., Rhee S.G.;
RT "Cloning, sequencing, and mutation of thiol-specific antioxidant gene
RT of Saccharomyces cerevisiae."
RL J. Biol. Chem. 268:16815-16821(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII.";
RL Nature 387:90-93(1997).
CC -!- FUNCTION: PHYSIOLOGICALLY IMPORTANT ANTIOXIDANT WHICH CONSTITUTES
CC AN ENZYMIC DEFENSE AGAINST SULFUR-CONTAINING RADICALS. CAN
CC PROVIDE PROTECTION AGAINST A THIOL-CONTAINING OXIDATION SYSTEM BUT
CC NOT AGAINST AN OXIDATION SYSTEM WITHOUT THIOL.
CC -!- SUBUNIT: HOMODIMER; DISULFIDE-LINKED, UPON OXIDATION.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- PTM: THE CYS-47-SH GROUP IS THE PRIMARY SITE OF OXIDATION BY
CC H(2)O(2), AND THE OXIDIZED CYS-47 (PROBABLY CYS-SOH) RAPIDLY
CC REACTS WITH CYS-170-SH OF THE OTHER SUBUNIT TO FORM AN
CC INTERMOLECULAR DISULFIDE. THIS DISULFIDE IS SUBSEQUENTLY REDUCED
CC BY THIOREDUXIN.

FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;
 Query Match 35.0%; Score 7; DB 1; Length 3033;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 12 LEDRDS 18
 |||||
 Db 658 LEDRDS 664
 RESULT 10
 ID_PEBB_HUMAN STANDARD; PRT; 182 AA.
 AC Q13951; Q13124;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Core-binding factor, beta subunit (CBF-beta)
 DE binding protein 2 beta subunit (PEBP2-beta) (SL3-3
 DE enhancer factor 1 beta subunit) (PEBP2-beta)
 DE enhancer factor 1 beta subunit (SL3/AKV core-binding factor beta
 DE subunit).
 GN CBFB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE OF 1-26 FROM N.A.
 RX MEDLINE=95331794; PubMed=7607682;
 RA Hajra A., Collins F.S.;
 RT "Structure of the leukemia-associated human CBFB gene.";
 RL Genomics 26:571-579 (1995).
 RN [2]
 RP ERRATUM.
 RX MEDLINE=97209219; PubMed=9064279;
 RA Hajra A., Collins F.S.;
 RL Genomics 38:107-107 (1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX TISSUE=EYE;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulláhy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A., Rodríguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [4]
 RP SEQUENCE OF 8-182 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93355287; PubMed=8351518;
 RA Liu P., Tarle S.A., Hajra A., Claxton D.F., Marlton P.,
 RA Freedman M., Siciliano M.J., Collins F.S.;

RT "Fusion between transcription factor CBF beta/PEBP2 beta and a myosin
 heavy chain in acute myeloid leukemia.";
 RL Science 261:1041-1044 (1993).
 RN [5]
 RP STRUCTURE BY NMR OF 4-141.
 RX MEDLINE=99332048; PubMed=10404215;
 RA Gogor M., Gupta V., Kim W.Y., Shigesada K., Ito Y., Werner M.H.;
 RT "Molecular insights into PEBP2/CBF beta-SMHC associated acute
 leukemia revealed from the structure of PEBP2/CBF beta.";
 RL Nat. Struct. Biol. 6:620-623 (1999).
 CC -!- FUNCTION: CBF BINDS TO THE CORE SITE, 5'-PYGPGGT-3', OF A NUMBER
 OF ENHANCERS AND PROMOTERS, INCLUDING MURINE LEUKEMIA VIRUS,
 CC POLYOMAVIRUS ENHANCER, T-CELL RECEPTOR ENHANCERS, LCK, IL-3 AND
 CC GM-CSF PROMOTERS. THE BETA SUBUNIT ACTS TO ENHANCE THE DNA BINDING
 OF THE ALPHA SUBUNIT.
 CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- DISEASE: A CHROMOSOMAL REARRANGEMENT, KNOWN AS PERICENTRIC
 INVERSION INV(16)(P13;Q22), PRODUCES A FUSION PROTEIN THAT
 CC CONSISTS OF THE 165 N-TERMINAL RESIDUES OF CBF-BETA (PEBP2) WITH
 THE TAIL REGION OF MYH11. THIS REARRANGEMENT IS ASSOCIATED WITH
 CC ACUTE MYELOID LEUKEMIA OF M4EO SUBTYPE.
 CC -!- SIMILARITY: BELONGS TO THE CBF-BETA FAMILY.
 CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
 WWW="http://www.infobiogen.fr/services/Chromancer/Genes/CBFBID45.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L20298; AAA02868.1; -;
 CC EMBL; BC018509; AAH18509.1; -;
 CC PIR; A56840; A56840.
 CC PIR; I59579; I59579.
 CC PDB; 1CL3; 01-JAN-00.
 CC TRANSFAC; T02259; -;
 CC Genew; HGNC:1539; CBFB.
 CC MIM; 121360; -;
 CC GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; TAS.
 CC GO; GO:0003700; P:transcription factor activity; TAS.
 CC GO; GO:0007048; P:oncogenesis; TAS.
 CC GO; GO:0006366; P:transcription from Pol II promoter; TAS.
 CC InterPro; IPR003417; CBF_beta.
 CC Pfam; PF02312; CBF_beta; 1.
 CC Nuclear protein; Chromosomal translocation; Proto-oncogene;
 CC 3D-structure. 165 166
 CC SITE BREAKPOINT FOR TRANSLOCATION TO FORM CBF
 CC SEQUENCE 182 AA; 21508 MW; 20FB1CC05FBFE4FB CRC64;
 BETA-MYH11 ONCOGENE IN AML, SUBTYPE M4EO.
 Query Match 30.0%; Score 6; DB 1; Length 182;
 Best Local Similarity 100.0%; Pred. No. 8;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 13 EDRDS 18
 |||||
 Db 154 EDRDS 159
 RESULT 11
 ID_PEBB_MOUSE STANDARD; PRT; 187 AA.
 AC Q08024; Q08025; Q62050; Q62051;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Core-binding factor, beta subunit (CBF-beta) (polyomavirus enhancer
 DE binding protein 2 beta subunit) (PEBP2-beta) (SL3-3
 DE enhancer factor 1 beta subunit) (SL3/AKV core-binding factor beta

FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 733 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 734 >737 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON TER 737 737
 SQ SEQUENCE 737 AA; 67DFAE11854122F2 CRC64;
 Query Match 35.0%; Score 7; DB 1; Length 737;
 Best Local Similarity 100.0%; Pred. No. 2.6;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 12 LEDRDRS 18
 |||||
 Db 658 LEDRDRS 664

RESULT 9
 POLG_HCVJ6 STANDARD; PRT; 3033 AA.
 AC P26660;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J6) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9204440; PubMed=1659196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 RA Machida A., Miyakawa Y., Mayumi M.;
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 RT from a human carrier: comparison with reported isolates for conserved
 RT and divergent regions".
 RL J. Gen. Virol. 72:2697-2704(1991).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D00944; BAA00792.1; --
 CC FIR; JQ1303; JQ1303.
 CC HSP; P27958; 1HEI.
 CC MEROPS; S29.001; --
 CC MEROPS; U39.001; --
 CC InterPro; IPR001410; DEAD.
 CC InterPro; IPR002522; HCV capsid.
 CC InterPro; IPR002531; HCV core.
 CC InterPro; IPR002519; HCV env.
 CC InterPro; IPR002531; HCV NS1.
 CC InterPro; IPR002518; HCV NS2.
 CC InterPro; IPR004109; HCV NS3.
 CC InterPro; IPR000745; HCV NS4a.
 CC InterPro; IPR001490; HCV NS4b.
 CC InterPro; IPR002868; HCV NS5a.
 CC InterPro; IPR002166; HCV RdRp.
 CC InterPro; IPR001650; Helicase C.
 CC InterPro; IPR007095; RNA pol Ds Ps.
 CC InterPro; IPR007094; RNA pol_PSVir.
 CC Pfam; PF01543; HCV_capsid; 1.
 CC Pfam; PF01542; HCV_core; 1.
 CC Pfam; PF01539; HCV env; 1.
 CC Pfam; PF01560; HCV NS1; 1.
 CC Pfam; PF01538; HCV NS2; 1.
 CC Pfam; PF02907; HCV NS3; 1.
 CC Pfam; PF01006; HCV NS4a; 1.
 CC Pfam; PF01001; HCV NS4b; 1.
 CC Pfam; PF01506; HCV NS5a; 1.
 CC Pfam; PF00271; Helicase C; 1.
 CC Pfam; PF00998; Viral RdRp; 1.
 CC ProDom; PD186062; HCV NS1; 1.
 CC SMART; SM00487; DEXDC; 1.
 CC Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 CC Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 CC REMOVED FROM CAPSID PROTEIN C BY THE
 CC CELLULAR AMINOPEPTIDASE.
 CC INIT_MET 1 1
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 477 477
 CC CARBOHYD 534 534
 CC CARBOHYD 542 542
 CC CARBOHYD 558 558
 CC CARBOHYD 578 578
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2018 3033
 CC CHAIN 347 369
 CC TRANSMEM 347 369
 CC ACT SITE 1087 1087
 CC ACT_SITE 1111 1111
 CC ACT_SITE 1169 1169
 CC NP_BIND 1234 1241
 CC SITE 1320 1323
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 4

QY 7 GRCOLEDRDR 17
 |||||
 Db 649 GRCOLEDRDR 659

RESULT 7

POLG_HCVJ5 STANDARD; PRT; 737 AA.

AC P27960;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
 DE protein (Envelope protein M); Major envelope protein E; Nonstructural
 DE proteins NS1 and NS2] (Fragment).
 OS Hepatitis C virus (isolate HC-J5) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_taxid=11112;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishihiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; D10075; BAA00969.1; -;
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 Pfam; PF01543; HCV_capsid; 1.
 Pfam; PF01542; HCV_core; 1.
 Pfam; PF01539; HCV_env; 1.
 Pfam; PF01560; HCV_NS1; 1.
 ProDom; PD186062; HCV_NS1; 1.
 KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 KW INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477

FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477

FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 737 737
 SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CRC64;

Query Match 35.0%; Score 7; DB 1; Length 737;

Best Local Similarity 100.0%; Pred. No. 2.6;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LEDRDRS 18

Db 658 LEDRDRS 664

RESULT 8

POLG_HCVJ7

ID POLG_HCVJ7 STANDARD; PRT; 737 AA.

AC P27961;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
 DE protein (Envelope protein M); Major envelope protein E; Nonstructural
 DE proteins NS1 and NS2] (Fragment).
 OS Hepatitis C virus (isolate HC-J7) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_taxid=11114;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishihiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; D10077; BAA00971.1; -;
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 Pfam; PF01543; HCV_capsid; 1.
 Pfam; PF01542; HCV_core; 1.
 Pfam; PF01539; HCV_env; 1.
 Pfam; PF01560; HCV_NS1; 1.
 ProDom; PD186062; HCV_NS1; 1.
 KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 KW INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477

FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477

REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 CAPSID PROTEIN C (POTENTIAL).
 MATRIX PROTEIN (POTENTIAL).

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX MEDLINE=9827846; PubMed=9568891;
RA Yan Y., Li Y., Munchi S., Sardana V., Cole J.L., Sardana M.,
RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.:
RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RL virus: a 2.2-A resolution structure in a hexagonal crystal form.";
RL Protein Sci. 7:837-847(1998).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC [RNA] (N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; MS9335; AAA72945.1; -.
DR PIR; A38465; GNVVTC.
DR PDB; 1A1Q; 25-MAR-98.
DR PDB; 1JXP; 14-JAN-98.
DR PDB; 1NS3; 08-APR-98.
DR PDB; 1C2P; 15-NOV-00.
DR PDB; 1GS5; 08-NOV-99.
DR PDB; 1GX5; 09-APR-02.
DR PDB; 1GX6; 10-APR-02.
DR PDB; 1QUV; 26-JUN-00.
DR PDB; 8OHM; 20-APR-99.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002538; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure. 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT INIT_MET CELLULAR AMINOPEPTIDASE.

FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2529 2529
FT CARBOHYD 2788 2788
FT STRAND 1031 1035
FT HELIX 1039 1047
FT STRAND 1050 1050
FT STRAND 1059 1063
FT STRAND 1068 1074
FT TURN 1075 1076
FT STRAND 1077 1081
FT HELIX 1082 1085
FT TURN 1086 1087
FT STRAND 1090 1092
FT TURN 1093 1094
FT STRAND 1095 1097
FT TURN 1101 1103
FT STRAND 1104 1107
FT STRAND 1108 1112
FT STRAND 1120 1120
FT STRAND 1122 1122
FT STRAND 1129 1133
FT TURN 1135 1136
FT STRAND 1139 1144
FT STRAND 1149 1157
FT HELIX 1158 1161
FT TURN 1162 1163
FT TURN 1165 1166
FT STRAND 1168 1171
FT TURN 1172 1174
FT STRAND 1175 1186
FT TURN 1187 1188
FT STRAND 1189 1197
FT HELIX 1198 1202
FT TURN 1203 1204
FT STRAND 1680 1688
SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCFFD9C CRC64;

Query Match 55.0%; Score 11; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 0.00043;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

$$\mathbb{Z}[\mathbf{f}]$$

KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydroxylase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2529 2529
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;
 Query Match 70.0%; Score 14; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 2.2e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 7 GERCDEDRDRSEL 20
 DB 649 GERCDEDRDRSEL 662
 RESULT 4
 POLG_HCV1 STANDARD; PRT; 3011 AA.
 AC P26664;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42; Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate 1) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11104;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=91172826; PubMed=1848704;
 RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
 RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
 RA Bradley D.W., Kuo G., Houghton M.,
 RT "Genetic organization and diversity of the hepatitis C virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA) (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M62321; AAA45676.1; -.
 CC PIR; A39166; GNWV03.
 CC PDB; 1AIV; 16-FEB-99.
 CC PDB; 1HET; 25-NOV-98.
 CC MEROPS; S29.001; -.
 CC MEROPS; U39.001; -.
 CC InterPro; IPR001410; DEAD.
 CC InterPro; IPR002522; HCV capsid.
 CC InterPro; IPR002521; HCV core.
 CC InterPro; IPR002519; HCV env.
 CC InterPro; IPR002531; HCV NS1.
 CC InterPro; IPR002518; HCV NS2.
 CC InterPro; IPR004109; HCV NS3.
 CC InterPro; IPR000745; HCV NS4a.
 CC InterPro; IPR001490; HCV NS4b.
 CC InterPro; IPR002868; HCV NS5a.
 CC InterPro; IPR002166; HCV RdRP.
 CC InterPro; IPR001650; Helicase C.
 CC InterPro; IPR007095; RNA_pol_DS_PS.
 CC InterPro; IPR007094; RNA_pol_PSVir.
 CC Pfam; PF01543; HCV capsid; 1.
 CC Pfam; PF01542; HCV core; 1.
 CC Pfam; PF01539; HCV env; 1.
 CC Pfam; PF01560; HCV NS1; 1.
 CC Pfam; PF01538; HCV NS2; 1.
 CC Pfam; PF02907; HCV NS3; 1.
 CC Pfam; PF01006; HCV NS4a; 1.
 CC Pfam; PF01001; HCV NS4b; 1.
 CC Pfam; PF01506; HCV NS5a; 1.
 CC Pfam; PF00271; Helicase C; 1.
 CC Pfam; PF00998; Viral RdRP; 1.
 CC ProDom; PD186062; HCV NS1; 1.
 CC SMART; SM00487; DEXDc1; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydroxylase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2529 2529
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;
 Query Match 70.0%; Score 14; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 2.2e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 7 GERCDEDRDRSEL 20
 DB 649 GERCDEDRDRSEL 662
 RESULT 4
 POLG_HCV1 STANDARD; PRT; 3011 AA.
 AC P26664;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42; Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate 1) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11104;
 RN [1]
 RP SEQUENCE FROM N.A.

DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECK BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 326575 MW; 94A1C77435D642BB CRC64;
 Query Match 70.0%; Score 14; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 2.2e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 GERCDELEDRSEL 20
 Db 649 GERCDELEDRSEL 662
 RESULT 3
 POLG_HCVTW STANDARD; PRT; 3010 AA.
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=31645;
 RX MEDLINE=92230206; PubMed=1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 RT "The Taiwanese hepatitis C virus genome: sequence determination and
 RT mapping the 5' terminus of viral genome and antigenomic RNA.";
 RL Virology 188:102-113(1992).
 CC !- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC !- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC !- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC !- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC !- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; M84754; -; NOT ANNOTATED_CDS.
 DR PIR; A40244; GNVVTV.
 DR PDB; 1N64; 25-FEB-03.
 DR PDB; 1NS3; 08-APR-98.
 DR MEROPS; S29.001; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRP.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.

PIR; A39253; GNWVCU.
 DR HSSP; P26663; LJXP.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR004109; HCV_NS4.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR007095; RNA_pol_Ds_ps.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; Helicase C; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;
 Query Match 70.0%; Score 14; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 2.2e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 GERCDDLRDRSEL 20
 DB 649 GERCDDLRDRSEL 662
 RESULT 2
 POLG_HCVJT STANDARD; PRT; 3010 AA.
 ID POLG_HCVJT
 AC Q00269;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP12) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-JT) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92295714; PubMed=1318627;
 RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 RT "Molecular cloning of hepatitis C virus genome from a single Japanese
 RT carrier: sequence variation within the same individual and among
 RT infected individuals.";
 RL Virus Res. 23:39-53(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA}(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D11168; BAA01943.1; -.
 CC PIR; A45573; A45573.
 CC PDB; 1A10; 25-MAR-98.
 CC PDB; 1JXP; 14-JAN-98.
 CC MEROPS; S29.001; -.
 CC MEROPS; U39.001; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR004109; HCV_NS4.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR007095; RNA_pol_Ds_ps.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; Helicase C; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;
 Query Match 70.0%; Score 14; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 2.2e-07;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:52 ; Search time 4.85 Seconds
(without alignments)
193.925 Million cell updates/sec

Title: US-09-973-025-91

Perfect score: 20

Sequence: 1 ACNWTGPERCDLEDRSEL 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	70.0	3010	1 POLG_HCVJA	P26662 h genome po
2	14	70.0	3010	1 POLG_HCVJT	Q00269 h genome po
3	14	70.0	3010	1 POLG_HCVTW	P29846 h genome po
4	14	70.0	3011	1 POLG_HCVH	P26664 h genome po
5	14	70.0	3011	1 POLG_HCVH	P27958 h genome po
6	11	55.0	3010	1 POLG_HCVBK	P26663 h genome po
7	7	35.0	737	1 POLG_HCVJ5	P27960 hepatitis c
8	7	35.0	737	1 POLG_HCVJ7	P27961 hepatitis c
9	7	35.0	3033	1 POLG_HCVJ6	P26660 h genome po
10	6	30.0	182	1 PEBB_MOUSE	Q13951 homo sapien
11	6	30.0	187	1 PEBB_MOUSE	Q08024 mus musculu
12	6	30.0	195	1 TSA1_YEAST	P34760 saccharomyc
13	6	30.0	195	1 TSA2_YEAST	Q04120 saccharomyc
14	6	30.0	393	1 C143_MYCTU	O53936 mycobacteri
15	6	30.0	473	1 POL_AVIRE	P03360 avian retic
16	6	30.0	483	1 MURE_CHLPN	Q92805 chlamydia p
17	6	30.0	542	1 CAB2_METKA	Q84207 methanopyru
18	6	30.0	606	1 SPB4_YEAST	P25808 saccharomyc
19	6	30.0	628	1 Y135_MYCTU	Q50598 mycobacteri
20	6	30.0	684	1 HELD_ECOLI	P15038 escherichia
21	6	30.0	686	1 TGM2_MOUSE	P211981 mus musculu
22	6	30.0	687	1 TGM2_BOVIN	P51176 bos taurus
23	6	30.0	687	1 TGM2_HUMAN	P21980 homo sapien
24	6	30.0	805	1 ACVS_STRCL	Q01757 streptomyce
25	6	30.0	843	1 VAV_RAT	P54100 rattus norv
26	6	30.0	845	1 VAV_HUMAN	P15498 homo sapien
27	6	30.0	845	1 VAV_MOUSE	P27870 mus musculu
28	6	30.0	870	1 YCS0_YEAST	P25623 saccharomyc
29	6	30.0	914	1 NRPI_CHICK	P79795 gallus gall
30	6	30.0	922	1 NRPI_RAT	Q9QWJ9 rattus norv
31	6	30.0	923	1 NRPI_HUMAN	O14786 homo sapien
32	6	30.0	923	1 NRPI_MOUSE	P97333 mus musculu
33	6	30.0	1043	1 CARY_BACST	Q92b63 bacillus st

34	6	30.0	1131	1 MOGL_CAEEL	P34498 caenorhabdi
35	6	30.0	1168	1 LMB3_MOUSE	Q61087 mus musculu
36	6	30.0	1172	1 LMB3_HUMAN	Q13751 homo sapien
37	6	30.0	1755	1 PEPL_MOUSE	Q9r269 mus musculu
38	6	30.0	1756	1 PEPL_HUMAN	O60437 homo sapien
39	6	30.0	1766	1 RPB1_TRYBB	P17545 trypanosoma
40	6	30.0	1766	1 RPB2_TRYBB	P17546 trypanosoma
41	6	30.0	3033	1 POLG_HCVJ8	P26661 h genome po
42	5	25.0	23	1 T2A_PARTE	Q27173 paramecium
43	5	25.0	23	1 T2B_PARTE	Q27174 paramecium
44	5	25.0	58	1 IHH_DEVMA	O13315 devario mal
45	5	25.0	60	1 HM09_CAEEL	P56407 caenorhabdi

ALIGNMENTS

RESULT 1
POLG_HCVJA STANDARD; PRT; 3010 AA.
AC P26662;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91088550; PubMed=2175903;
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA Sugimura T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from
RT Japanese patients with non-A, non-B hepatitis.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=91192160; PubMed=1849488;
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K.,
RA Ohkoshi S., Shimotohno K.;
RT "Molecular structure of the Japanese hepatitis C viral genome.";
RL FEBS Lett. 280:325-328(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D90208; BAA14233.1; -

THIS PAGE BLANK (USPTO)

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LEDRDRSEL 20
 |||||
 Db 654 LEDRDRSEL 662

RESULT 13

S18030
 genome polyprotein - hepatitis C virus (isolate JK1)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4b; nonstructural protein NS5b; nonstructural protein NS5)
 C:Species: hepatitis C virus
 A:Variety: isolate JK1
 C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient
 A:Reference number: S18028
 A:Accession: S18030
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <HOW>
 A:Cross-references: EMBL:X61596; NID:G59478; PIDN:CA443793.1; PID:G59479
 A:Experimental source: isolate JK1 from an individual
 R:Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.
 Arch. Virol. 128, 163-169, 1993
 A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolated
 A:Reference number: A48332; MUID:93119270; PMID:8380322
 A:Accession: S33570
 A:Molecule type: genomic RNA
 A:Residues: 1-547, 'T', '549-621', 'V', '623-624', 'S', '626-652', 'DL', '655-761', 'T', '763-782' <HOW>
 A:Cross-references: EMBL:X61591
 A:Note: this sequence is inconsistent with the nucleotide translation
 A:Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320 as Trp, and TTC for residue 771 as Ser
 A:Note: sequence extracted from NCBI backbone (NCBI:121747, NCBI:121748)
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serine
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (Asn)

Query Match 40.0%; Score 8; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 EDRDRSEL 20
 |||||
 Db 655 EDRDRSEL 662

RESULT 14

AF2371
 hypothetical protein alr4526 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AF2371
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuchihara, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AF2371

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-447 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA76225.1; PID:gl7133662; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr4526

Query Match 35.0%; Score 7; DB 2; Length 447;
 Best Local Similarity 100.0%; Pred. No. 4.6;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 DRDRSEL 20
 |||||

Db 152 DRDRSEL 158

RESULT 15

JQ1366

polyprotein - hepatitis C virus (French isolate) (fragments)

C:Species: hepatitis C virus

C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000

C:Accession: JQ1366

R:Kremsdorf, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.

J. Gen. Virol. 72, 2557-2561, 1991

A:Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication

A:Reference number: JQ1366; MUID:92013977; PMID:1655961

A:Accession: JQ1366

A:Molecule type: genomic RNA

A:Residues: 1-716 <KRE>

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: glycoprotein; polypeptide

F:84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #sta

Query Match 35.0%; Score 7; DB 2; Length 716;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 DRDRSEL 20
 |||||

Db 323 DRDRSEL 329

Search completed: November 21, 2003, 21:11:39

Job time : 9.25 secs

RESULT 9
GNWVC

genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
Accession: A38465

R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J. Virol. 65, 1105-1113, 1991
A:Title: Structure and organization of the hepatitis C virus genome isolated from human
A:Reference number: A38465; MUID:91140698; PMID:1847440

A:Accession: A38465

A:Molecule type: genomic RNA

A:Residues: 1-3010 <TAK>

A:Cross-references: EMBL:M58335; NID:g329770; PIDN:AAA72945.1; PID:g329771

C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitis C virus genome polyprotein

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match

Best Local Similarity 55.0%; Score 11; DB 1; Length 3010;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCDELDNR 17

|||||

DB 649 GERCDELDNR 659

RESULT 10

JC5620

genome polyprotein - hepatitis C virus (isolate EUH1480)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus

C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001

C:Accession: JC5620

R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.

Biochem. Biophys. Res. Commun. 236, 44-49, 1997

A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant

A:Reference number: JC5620; MUID:97366593; PMID:9223423

A:Accession: JC5620

A:Molecule type: mRNA

A:Residues: 1-3014 <CHA>

A:Cross-references: GB:Y13184

A:Experimental source: genotype 5a, which predominates in South Africa

A:Note: the translation of the nucleotide sequence is not complete in this paper

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin

F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: capsid protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:384-408/Region: hypervariable #status predicted

F:730-730/Product: nonstructural protein NS1 #status predicted <NS1>

F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>

F:1008-1616/Product: hepatitis C virus genome polyprotein

F:1231-1238/Region: nucleotide-binding motif A (P-loop)

F:1313-1318/Region: nucleotide-binding motif B

F:1317-1320/Region: DEXH motif

F:1617-1863/Product: nonstructural protein NS4a #status predicted <N4A>
F:1864-2014/Product: nonstructural protein NS4b #status predicted <N4B>
F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F:2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 55.0%; Score 11; DB 1; Length 3014;

Best Local Similarity 100.0%; Pred. No. 0.0014;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCDELDNR 17

|||||

DB 650 GERCDELDNR 660

RESULT 11

S18031

genome polyprotein - hepatitis C virus (isolate JK2) (fragment)

N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein

C:Species: hepatitis C virus

A:Variety: isolate JK2

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000

C:Accession: S18031

R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.

submitted to the EMBL Data Library, September 1991

A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus is

A:Reference number: S18029

A:Accession: S18031

A:Molecule type: genomic RNA

A:Residues: 1-782 <HON>

A:Cross-references: EMBL:X61593

A:Experimental source: isolate JK2

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural

F:1-191/Product: core protein #status predicted <MAT1>

F:192-383/Product: envelope protein 1 #status predicted <MAT2>

F:384-733/Product: NS1/E2 protein #status predicted <MAT3>

F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match

Best Local Similarity 45.0%; Score 9; DB 2; Length 782;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LEDRDRSEL 20

|||||

DB 654 LEDRDRSEL 662

RESULT 12

S18032

genome polyprotein - hepatitis C virus (isolate JK4) (fragment)

N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein

C:Species: hepatitis C virus

A:Variety: isolate JK4

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000

C:Accession: S18032

R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.

submitted to the EMBL Data Library, September 1991

A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus is

A:Reference number: S18029

A:Accession: S18032

A:Molecule type: genomic RNA

A:Residues: 1-782 <HON>

A:Cross-references: EMBL:X61594

A:Experimental source: isolate JK4

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural

F:1-191/Product: core protein #status predicted <MAT1>

F:192-383/Product: envelope protein 1 #status predicted <MAT2>

F:384-733/Product: NS1/E2 protein #status predicted <MAT3>

F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match

Best Local Similarity 45.0%; Score 9; DB 2; Length 782;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C/Accession: S40770; PC1285
R/Okamoto, H.
submitted to the EMBL Data Library, March 1992
A/Reference number: S40770
A/Accession: S40770
A/Molecule type: genomic RNA
A/Residues: 1-3011 <OKA>
A/Cross-references: EMBL:D10749; NID:G221586; PIDN:BAA01582.1; PID:G221587
R/Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa
Jpn. J. Exp. Med. 60, 167-177, 1990
A/Title: The 5'-terminal sequence of the hepatitis C virus genome.
A/Reference number: PC1284; MUID:91013116; PMID:2170712
A/Accession: PC1285
A/Molecule type: genomic RNA
A/Residues: 1-513 <OK2>
A/Cross-references: GB:D00831; NID:G221511; PIDN:BAA00705.1; PID:G221512
A/Experimental source: isolate HC-J1
C/Superfamily: hepatitis C virus genome polyprotein
C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; poly
F/2-115/Product: capsid protein C #status predicted <CPC>
F/116-191/Product: envelope protein E #status predicted <EPM>
F/192-389/Product: major envelope protein E #status predicted <MEE>
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F/1007-1615/Product: hepatitis C virus polyprotein #status predicted <NS3>
F/1230-1237/Region: nucleotide-binding motif A (P-loop)
F/1312-1317/Region: nucleotide-binding motif B
F/1316-1319/Region: DEXH motif
F/1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F/1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F/2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 70.0%; Score 14; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 9e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps

Qy 7 GRCLEDRDSEL 20
|||||
Db 649 GRCLEDRDSEL 662

RESULT 8
S19876
genome polyprotein - hepatitis C virus (isolate JK5) (fragment)
N/Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/
C/Species: hepatitis C virus
A/Variety: isolate JK5
C/Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C/Accession: S19876
R/Honda, M.; Kaneo, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A/Description: Sequence analysis of putative structural regions of Hepatitis
A/Reference number: S18029
A/Accession: S19876
A/Molecule type: genomic RNA
A/Residues: 1-782 <HON>
A/Cross-references: EMBL:X61595; NID:G59486; PIDN:CAA43792.1; PID:G59487
A/Experimental source: isolate JK5
C/Superfamily: hepatitis C virus genome polyprotein
C/Keywords: capsid protein; core protein; envelope protein; glycoprotein; n
F/1-191/Product: core protein #status predicted <MAT1>
F/192-383/Product: envelope protein 1 #status predicted <MAT2>
F/384-733/Product: NS1/E2 protein #status predicted <MAT3>
F/734-782/Product: nonstructural protein 2 (fragment) #status predicted <MA

Query Match 60.0%; Score 12; DB 2; Length 782;
Best Local Similarity 100.0%; Pred. No. 3.6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps

Qy 9 RCDLEDRDSEL 20
|||||
Db 651 RCDLEDRDSEL 662

C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40244
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992
A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
A:Reference number: A40244; MOID:92230206; PMID:131449
A:Accession: A40244
A:Molecule type: genomic RNA
A:Residues: 1-3010 <CHE>
A:Cross-references: GB:M84754
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructur
F:1-15/Product: capsid protein C #status predicted <PC>

F.192-389/Product: major envelope protein E #status predicted <MSE>
F.390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F.7730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F.1007-1615/Product: hepatitis B virus surface antigen #status predicted <NS3>
F.1230-1237/Region: nucleotide-binding motif A (P-loop)
F.1312-1317/Region: nucleotide-binding motif B
F.13116-1319/Region: DEXH motif

```

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS>
F:196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,20
Query Match 70.0%; Score 14; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 9e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GERCLEDRDRSEL 20
    |||||
Db 649 GERCLEDRDRSEL 662

RESULT 5
GNWVC3
genome polyprotein - hepatitis C virus (strain HCV-1)
N:Contains: capsid protein C; envelope protein M; hepatitisvirin (EC 3.4.21.98) (nonstruct

```

A:Contains capsid protein, nonstructural protein NS4b; nonstructural protein NS5
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 19-Jan-2001
 C:Accession: A39166; PQ0403; PQ0404
 R:Choo, Q.L.; Rishman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Co
 Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
 A:Title: Genetic organization and diversity of the hepatitis C virus.
 A:Reference number: A39166; MUID:91172826; PMID:1848704
 A:Accession: A39166
 A:Molecule type: mRNA
 A:Residues: 1-3011 <GB>
 A:Cross-references: CHM62321; NID:G329873; PIDN:AAA45676.1; PID:G329874
 R:Chan, S.W.; McMahon, P.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.
 J. Gen. Virol. 73, 1131-1141, 1992
 A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to
 A:Reference number: PQ0393; MUID:92268871; PMID:1316939
 A:Accession: PQ0403

A.Molecule type: genomic RNA
A.Residues: 1577-1633 <CH>
A.Cross-references: DDBJ:D10128
A.Experimental source: isolates E-b16
A.Accession: PQ0404
A.Status: preliminary
A.Molecule type: genomic RNA
A.Residues: 1577-1633 <CH2>

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NSI>

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:46:27 ; Search time 9.25 Seconds
(without alignments)
207.932 Million cell updates/sec

Title: US-09-973-025-91

Perfect score: 20

Sequence: 1 ACNWTPGERCDLEDRDRSEL 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 76:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	70.0	787	2	GNWVCJ
2	14	70.0	3010	1	GNWVCJ
3	14	70.0	3010	1	A45573
4	14	70.0	3010	1	GNWVTW
5	14	70.0	3011	1	GNWVC3
6	14	70.0	3011	1	GNWVCH
7	14	70.0	3011	1	S40770
8	12	60.0	782	2	S19876
9	11	55.0	3010	1	GNWVC
10	11	55.0	3014	1	JCS620
11	9	45.0	782	2	S18031
12	9	45.0	782	2	S18032
13	8	40.0	3010	1	S18030
14	7	35.0	447	2	AF2371
15	7	35.0	716	2	JQ1366
16	7	35.0	3033	1	JQ1303
17	6	30.0	128	2	AE0528
18	6	30.0	148	2	C48124
19	6	30.0	175	2	I59579
20	6	30.0	182	2	A46107
21	6	30.0	187	2	B48124
22	6	30.0	196	2	A47362
23	6	30.0	196	2	S69732
24	6	30.0	293	2	B38169
25	6	30.0	336	2	T44988
26	6	30.0	342	2	T27348
27	6	30.0	349	2	S68092
28	6	30.0	393	1	C70929
29	6	30.0	473	2	A03959

30	6	30.0	483	2	H86542
31	6	30.0	483	2	D72080
32	6	30.0	500	2	B72522
33	6	30.0	548	2	A41302
34	6	30.0	585	2	T37034
35	6	30.0	606	2	S14942
36	6	30.0	617	2	T02121
37	6	30.0	628	2	D70722
38	6	30.0	669	2	F84433
39	6	30.0	684	1	HJECDA
40	6	30.0	684	2	F90759
41	6	30.0	684	2	D85623
42	6	30.0	687	1	A39045
43	6	30.0	687	1	B39045
44	6	30.0	687	2	S19680
45	6	30.0	698	2	C84836

ALIGNMENTS

RESULT 1

PN0677 hypothetical protein 787 - hepatitis C virus (fragment)

C:Species: hepatitis C virus

C>Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000

C:Accession: PN0677

R:Cho, S.H.; Yoon, J.I.; Chang, J.E.; Ahn, B.M.; Lee, C.H.; Lee, Y.I.

Biochem. Biophys. Res. Commun. 196, 780-788, 1993

A:Title: Genomic typing of hepatitis C viruses from Korean patients: Implications of ge

C:Keywords: glycoprotein; nonstructural protein

A:Reference number: PN0677; MUID:94059104; PMID:8240354

A:Accession: PN0677

A:Molecule type: mRNA

A:Residues: 1-787 <CHO>

A:Cross-references: GB:L20498; NID:gl381031; PIDN:AAB02608.1; PID:gl381032

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: glycoprotein; nonstructural protein

F:196,209,234,250,305,325,421,427,452,536,544,560,580,627,649/Binding site: carbohydrate

Query Match 70.0%; Score 14; DB 2; Length 787;

Best Local Similarity 100.0%; Pred. No. 2.7e-07;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20

DB 653 GERCLEDRDRSEL 666

RESULT 2

GNWVCJ

genome polyprotein - hepatitis C virus (strain J)

N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructur

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001

C:Accession: A39253; PS0086

R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Okoshi, S.; Sugimura, T.; Shimo

Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990

A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients

A:Reference number: A39253; MUID:91088550; PMID:2175903

A:Accession: A39253

A:Molecule type: genomic RNA

A:Residues: 1-3010 <KAT>

A:Cross-references: GB:D90208; NID:g221610; PIDN:BAAL4233.1; PID:g221611

R:Kato, N.; Okoshi, S.; Shimotohno, K.

Proc. Jpn. Acad. 65B, 219-223, 1989

A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence var

A:Reference number: PS0086

A:Accession: PS0086

A:Molecule type: genomic RNA

A:Residues: 2650-2707 <KA2>

A:Experimental source: Japanese isolate

C:Comment: The cleavage sites of this polyprotein have not been determined.

Db 649 GERCDLEDRDR 659

RESULT 12

US-09-350-841A-1085
; Sequence 1085, Application US/09350841A
; GENERAL INFORMATION:
; APPLICANT: Jeffs, Peter;
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION OF MEMBRANE
; FILE REFERENCE: 7872-066-999
; CURRENT APPLICATION NUMBER: US/09/350,841A
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 1946
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1085
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-09-350-841A-1085

Query Match 50.0%; Score 10; DB 5; Length 28;
Best Local Similarity 100.0%; Pred. No. 0.00078;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 DLEDRDRSEL 20
Db 17 DLEDRDRSEL 26

RESULT 13

US-10-351-641-1085
; Sequence 1085, Application US/10351641
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
; APPLICANT: Lambert, D.
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; FILE REFERENCE: 7872-100
; CURRENT APPLICATION NUMBER: US/10/351,641
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: 09/350,641
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/315,304
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1757
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1085
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-10-351-641-1085

Query Match 50.0%; Score 10; DB 6; Length 28;
Best Local Similarity 100.0%; Pred. No. 0.00078;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 DLEDRDRSEL 20
Db 17 DLEDRDRSEL 26

RESULT 14

US-09-350-369D-1704

; Sequence 1704, Application US/09350369D
; GENERAL INFORMATION:
; APPLICANT: Samuel E. Hopkins
; APPLICANT: Elizabeth M. Dimassimo
; APPLICANT: Pamela P. Rusnak
; APPLICANT: Thomas Michael Venetta
; TITLE OF INVENTION: Methods and Compositions for Detection of HIV GP41 Peptides
; FILE REFERENCE: 7872-068-999
; CURRENT APPLICATION NUMBER: US/09/350,369D
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 2088
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1704
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus-1 (HIV-1)
US-09-350-369D-1704

Query Match 50.0%; Score 10; DB 5; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.00082;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 DLEDRDRSEL 20
Db 19 DLEDRDRSEL 28

RESULT 15

US-09-350-369E-1704
; Sequence 1704, Application US/09350369E
; GENERAL INFORMATION:
; APPLICANT: Samuel E. Hopkins
; APPLICANT: Elizabeth M. Dimassimo
; APPLICANT: Pamela P. Rusnak
; APPLICANT: Thomas Michael Venetta
; TITLE OF INVENTION: Methods and Compositions for Detection of HIV GP41 Peptides
; FILE REFERENCE: 7872-068-999
; CURRENT APPLICATION NUMBER: US/09/350,369E
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 2088
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1704
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus-1 (HIV-1)
US-09-350-369E-1704

Query Match 50.0%; Score 10; DB 5; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.00082;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 DLEDRDRSEL 20
Db 19 DLEDRDRSEL 28

Search completed: November 21, 2003, 22:12:59
Job time : 8.55 secs

; Sequence 211, Application PC/TUS0320322
; GENERAL INFORMATION:
; APPLICANT: Dana-Farber Cancer Institute, Inc., et al.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
; FILE REFERENCE: DFN-043PC
; CURRENT APPLICATION NUMBER: PCT/US03/20322
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 211
; LENGTH: 2280
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
PCT-US03-20322-211

Query Match 70.0%; Score 14; DB 1; Length 2280;
Best Local Similarity 100.0%; Pred. No. 2e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
Db 649 GERCLEDRDRSEL 662

RESULT 8

PCT-US03-19834-2
; Sequence 2, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; TITLE OF INVENTION: METHODS FOR TREATING HYPERLIPIDEMIA
; FILE REFERENCE: IOWA:045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-2

Query Match 70.0%; Score 14; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.4e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
Db 649 GERCLEDRDRSEL 662

RESULT 9

US-10-296-734-406
; Sequence 406, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la consensus polyprotein
US-10-296-734-406

Query Match 70.0%; Score 14; DB 6; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.4e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
Db 649 GERCLEDRDRSEL 662

RESULT 10

US-10-296-734-810
; Sequence 810, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 810
; LENGTH: 5985
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC la savine
US-10-296-734-810

Query Match 70.0%; Score 14; DB 6; Length 5985;
Best Local Similarity 100.0%; Pred. No. 3.9e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
Db 5406 GERCLEDRDRSEL 5419

RESULT 11

PCT-US03-20409-3
; Sequence 3, Application PC/TUS0320409
; GENERAL INFORMATION:
; APPLICANT: Merck & Co., Inc.
; TITLE OF INVENTION: HEPATITIS C VIRUS PARTICLE FORMATION
; FILE REFERENCE: 21123 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/20409
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/393,167
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1026
; TYPE: PRT
; ORGANISM: HCV
PCT-US03-20409-3

Query Match 55.0%; Score 11; DB 1; Length 1026;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDR 17
|||||

```
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-685-435-31

Query Match          70.0%; Score 14; DB 6; Length 33;
Best Local Similarity 100.0%; Pred. No. 9.6e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDLDRSEL 20
Db 9 GERCLEDLDRSEL 22

RESULT 3
PCT-US03-19834-3
; Sequence 3, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; FILE REFERENCE: IOWA.045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-3

Query Match          70.0%; Score 14; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDLDRSEL 20
Db 267 GERCLEDLDRSEL 280

RESULT 4
US-10-655-562-4
; Sequence 4, Application US/10655562
; GENERAL INFORMATION:
; APPLICANT: BISHOP-HURLEY, SHARON L.
; APPLICANT: SCHMIDT, FRANCIS J.
; APPLICANT: SMITH, ARNOLD L.
; TITLE OF INVENTION: PHAGE-DISPLAY PEPTIDES AS NOVEL ANTIMICROBIAL AGENTS
; FILE REFERENCE: UWO.022US
; CURRENT APPLICATION NUMBER: US/10/655,562
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: 60,409,909
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-655-562-4

Query Match          70.0%; Score 14; DB 6; Length 333;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDLDRSEL 20
Db 1416 GERCLEDLDRSEL 1429

RESULT 5
PCT-US03-33610-4
; Sequence 4, Application PC/TUS0333610
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: SELBY, Mark
; APPLICANT: PALLIARD, Xavier
; TITLE OF INVENTION: ACTIVATION OF HCV-SPECIFIC T-CELLS
; FILE REFERENCE: 2300-1612.60
; CURRENT APPLICATION NUMBER: PCT/US03/33610
; CURRENT FILING DATE: 2003-10-24
; PRIOR APPLICATION NUMBER: 10/281,341
; PRIOR FILING DATE: 2002-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
PCT-US03-33610-4

Query Match          70.0%; Score 14; DB 1; Length 637;
Best Local Similarity 100.0%; Pred. No. 7.9e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDLDRSEL 20
Db 477 GERCLEDLDRSEL 490

RESULT 6
US-10-296-734-816
; Sequence 816, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 816
; LENGTH: 1997
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette C
US-10-296-734-816

Query Match          70.0%; Score 14; DB 6; Length 1997;
Best Local Similarity 100.0%; Pred. No. 1.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDLDRSEL 20
Db 1416 GERCLEDLDRSEL 1429

RESULT 7
PCT-US03-20322-211
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.55 Seconds
(without alignments)
117.536 Million cell updates/sec

Title: US-09-973-025-91
Perfect score: 20
Sequence: 1 ACNWTGRCDEDRSEL 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0
Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_AA_New.*
1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	70.0	30	6	US-10-296-734-494
2	14	70.0	33	6	US-10-685-435-31
3	14	70.0	333	1	PCT-US03-19834-3
4	14	70.0	333	6	US-10-655-562-4
5	14	70.0	637	1	PCT-US03-33610-4
6	14	70.0	1997	6	US-10-296-734-816
7	14	70.0	2280	1	PCT-US03-20322-211
8	14	70.0	3011	1	PCT-US03-19834-2
9	14	70.0	3011	6	US-10-296-734-406
10	14	70.0	5985	6	US-10-296-734-810
11	11	55.0	1026	1	PCT-US03-20409-3
12	10	50.0	28	5	US-09-350-841A-1085
13	10	50.0	28	6	US-10-351-641-1085
14	10	50.0	30	5	US-09-350-369D-1704
15	10	50.0	30	5	US-09-350-369E-1704
16	10	50.0	30	6	US-10-296-734-492
17	10	50.0	32	5	US-09-350-841A-1217
18	10	50.0	32	6	US-10-351-641-1217
19	10	50.0	34	5	US-09-350-841A-1904
20	10	50.0	34	5	US-09-350-369E-1828
21	10	50.0	34	5	US-09-350-369E-1828
22	10	50.0	34	6	US-10-351-641-1715
23	10	50.0	2011	6	US-10-296-734-812
24	9	45.0	9	6	US-10-440-390-39
25	8	40.0	347	6	US-10-664-391-9
26	8	40.0	539	6	US-10-664-391-11

27	7	35.0	3033	6	US-10-009-002-5	Sequence 5, Appli
28	6	30.0	349	7	US-60-493-369-178	Sequence 178, App
29	6	30.0	349	7	US-60-502-656-76	Sequence 76, Appl
30	6	30.0	349	7	US-60-513-690-229	Sequence 229, App
31	6	30.0	380	6	US-10-048-687A-4	Sequence 4, Appli
32	6	30.0	418	1	PCT-US03-18787-118	Sequence 118, App
33	6	30.0	418	1	PCT-US03-19069-118	Sequence 118, App
34	6	30.0	548	7	US-60-493-369-179	Sequence 179, App
35	6	30.0	548	7	US-60-502-656-78	Sequence 78, Appl
36	6	30.0	548	7	US-60-512-690-230	Sequence 230, App
37	6	30.0	678	1	PCT-US02-24483-22	Sequence 22, Appl
38	6	30.0	684	6	US-10-380-533-136	Sequence 136, App
39	6	30.0	687	7	US-60-493-369-176	Sequence 176, App
40	6	30.0	687	7	US-60-493-369-177	Sequence 177, App
41	6	30.0	687	7	US-60-493-369-180	Sequence 180, App
42	6	30.0	687	7	US-60-493-369-181	Sequence 181, App
43	6	30.0	687	7	US-60-502-656-75	Sequence 75, Appl
44	6	30.0	687	7	US-60-502-656-77	Sequence 77, Appl
45	6	30.0	687	7	US-60-502-656-79	Sequence 79, Appl

ALIGNMENTS

RESULT 1
US-10-296-734-494
; Sequence 494, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 494
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 44
US-10-296-734-494

Query Match	70.0%	Score 14;	DB 6;	Length 30;
Best Local Similarity	100.0%	Pred. No. 9e-08;		
Matches 14;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	7	GERCDLEDRDRSEL 20		
Db	6	GERCDLEDRDRSEL 19		
RESULT 2				
US-10-685-435-31				
; Sequence 31, Application US/10685435				
; GENERAL INFORMATION:				
; APPLICANT: Innogenetics N.V.				
; TITLE OF INVENTION: Multi-mer peptides derived from Hepatitis C Virus				
; TITLE OF INVENTION: envelope proteins for diagnostic use and vaccination				
; FILE REFERENCE: PCT98.75.HCV30				
; CURRENT APPLICATION NUMBER: US/10/685,435				
; CURRENT FILING DATE: 2003-10-16				
; PRIOR APPLICATION NUMBER: US/09/566,266B				
; PRIOR FILING DATE: 2000-05-05				
; PRIOR APPLICATION NUMBER: 97870179.5				
; PRIOR FILING DATE: 1997-11-06				
; NUMBER OF SEQ ID NOS: 41				
; SOFTWARE: patenting Ver. 2.1				
; SEQ ID NO 31				

APPLICATION NUMBER: 08/262,037
FILING DATE: 07-June-1995
APPLICATION NUMBER: 07/719,819
FILING DATE: 24-June-1991
APPLICATION NUMBER: 07/667,275
FILING DATE: 11-Mar-1991
APPLICATION NUMBER: 07/651,735
FILING DATE: 07-Feb-1991
APPLICATION NUMBER: 07/558,799
FILING DATE: 26-July-1990
APPLICATION NUMBER: 07/510,153
FILING DATE: 16-April-1990
ATTORNEY/AGENT INFORMATION:
NAME: Maria C. H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4043 US7
TELEPHONE: 212-758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: Amino acid
STRANDEDNESS:
TOPOLOGY: Unknown
US-08-477-582-45

Query Match 70.0%; Score 14; DB 8; Length 23;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GERCDEDRDRSEL 20
Db 9 GERCDEDRDRSEL 22

RESULT 15
US-08-480-253-45
Sequence 45, Application US/08480253
GENERAL INFORMATION:
APPLICANT: Chang Yi Wang and Barbara Hosein
TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
NUMBER OF SEQUENCES: 136
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVE.
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,253
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/262,037
FILING DATE: 07-June-1995
APPLICATION NUMBER: 07/719,819
FILING DATE: 24-June-1991
APPLICATION NUMBER: 07/667,275
FILING DATE: 11-Mar-1991
APPLICATION NUMBER: 07/651,735
FILING DATE: 07-Feb-1991
APPLICATION NUMBER: 07/558,799
FILING DATE: 26-July-1990

APPLICATION NUMBER: 07/510,153
FILING DATE: 16-April-1990
ATTORNEY/AGENT INFORMATION:
NAME: Maria C. H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4043 US9
TELEPHONE: 212-758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: Amino acid
STRANDEDNESS:
TOPOLOGY: Unknown
US-08-480-253-45

Query Match 70.0%; Score 14; DB 8; Length 23;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GERCDEDRDRSEL 20
Db 9 GERCDEDRDRSEL 22

Search completed: November 21, 2003, 22:09:54
Job time : 163.1 secs

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477.072
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/262.037
FILING DATE: 17-June-1994
APPLICATION NUMBER: 07/719.819
FILING DATE: 24-June-1991
APPLICATION NUMBER: 07/667.275
FILING DATE: 11-Mar-1991
APPLICATION NUMBER: 07/651.735
FILING DATE: 07-Feb-1991
APPLICATION NUMBER: 07/558.799
FILING DATE: 26-July-1990
APPLICATION NUMBER: 07/510.153
FILING DATE: 16-April-1990
ATTORNEY/AGENT INFORMATION:
NAME: Maria C. H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4043 US8
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: Amino acid
STRANDEDNESS:
TOPOLOGY: Unknown
US-08-477-072-45

Query Match 70.0%; Score 14; DB 8; Length 23;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 7 GERCLEDRDRSEL 20
|||||
Db 9 GERCLEDRDRSEL 22

RESULT 14
US-08-477-582-45
Sequence 45, Application US/08477582
GENERAL INFORMATION:
APPLICANT: Chang Yi Wang and Barbara Hosein
TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
NUMBER OF SEQUENCES: 136
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVE.
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477.582
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:

ADDRESSEE: ABBOTT LABORATORIES CHAD377/AP6D
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: ILLINOIS
COUNTRY: U.S.A.
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/507,740
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA: US/08/373,920
FILING DATE:
APPLICATION NUMBER: US/08/183,207
FILING DATE:
APPLICATION NUMBER: US 07/760,292
FILING DATE: 16-SEP-1991
APPLICATION NUMBER: US 07/456,162
FILING DATE: 22-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/610,180
FILING DATE: 07-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: POREMBSKIP, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 4767.US.P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-507-740-20
Query Match 70.0%; Score 14; DB 9; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 GERCDEDRDRSEL 20
DB 7 GERCDEDRDRSEL 20
RESULT 10
US-08-905-054-20
Sequence 20, Application US/08905054
GENERAL INFORMATION:
APPLICANT: Abbott Laboratories
APPLICANT: Lesniewski, Richard R.
APPLICANT: Leung, Tat K.
TITLE OF INVENTION: HEPATITIS C ASSAY
FILE REFERENCE: 4767.US.C7
CURRENT APPLICATION NUMBER: US/08/905,054
CURRENT FILING DATE: 1997-08-01
PRIOR APPLICATION NUMBER: US 08/707,355
PRIOR FILING DATE: 1996-09-04
PRIOR APPLICATION NUMBER: US 08/507,740
PRIOR FILING DATE: 1995-07-26
PRIOR APPLICATION NUMBER: US 08/373,920
PRIOR FILING DATE: 1995-01-17
PRIOR APPLICATION NUMBER: US 08/183,207
PRIOR FILING DATE: 1994-01-18
PRIOR APPLICATION NUMBER: US 07/760,292
PRIOR FILING DATE: 1991-09-16
PRIOR APPLICATION NUMBER: US 07/610,180
PRIOR FILING DATE: 1990-11-07
PRIOR APPLICATION NUMBER: US 07/456,162
PRIOR FILING DATE: 1989-12-22
NUMBER OF SEQ ID NOS: 30
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 20
LENGTH: 22
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Modified Peptides
US-08-905-054B-20
Query Match 70.0%; Score 14; DB 13; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 GERCDEDRDRSEL 20
DB 7 GERCDEDRDRSEL 20
RESULT 12
US-08-475-482-45
Sequence 45, Application US/08475482
GENERAL INFORMATION:
APPLICANT: Chang Yi Wang and Barbara Hosein
TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR

PRIOR APPLICATION NUMBER: US 07/610,180
PRIOR FILING DATE: 1990-11-07
PRIOR APPLICATION NUMBER: US 07/456,162
PRIOR FILING DATE: 1989-12-22
NUMBER OF SEQ ID NOS: 30
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 20
LENGTH: 22
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Modified Peptides
US-08-905-054-20
Query Match 70.0%; Score 14; DB 13; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 GERCDEDRDRSEL 20
DB 7 GERCDEDRDRSEL 20
RESULT 11
US-08-905-054B-20
Sequence 20, Application US/08905054B
GENERAL INFORMATION:
APPLICANT: Abbott Laboratories
APPLICANT: Lesniewski, Richard R.
APPLICANT: Leung, Tat K.
TITLE OF INVENTION: HEPATITIS C ASSAY
FILE REFERENCE: 4767.US.C7
CURRENT APPLICATION NUMBER: US/08/905,054B
CURRENT FILING DATE: 1997-08-01
PRIOR APPLICATION NUMBER: US 08/707,355
PRIOR FILING DATE: 1996-09-04
PRIOR APPLICATION NUMBER: US 08/507,740
PRIOR FILING DATE: 1995-07-26
PRIOR APPLICATION NUMBER: US 08/373,920
PRIOR FILING DATE: 1995-01-17
PRIOR APPLICATION NUMBER: US 08/183,207
PRIOR FILING DATE: 1994-01-18
PRIOR APPLICATION NUMBER: US 07/760,292
PRIOR FILING DATE: 1991-09-16
PRIOR APPLICATION NUMBER: US 07/610,180
PRIOR FILING DATE: 1990-11-07
PRIOR APPLICATION NUMBER: US 07/456,162
PRIOR FILING DATE: 1989-12-22
NUMBER OF SEQ ID NOS: 30
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 20
LENGTH: 22
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Modified Peptides
US-08-905-054B-20
Query Match 70.0%; Score 14; DB 13; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 GERCDEDRDRSEL 20
DB 7 GERCDEDRDRSEL 20
RESULT 12
US-08-475-482-45
Sequence 45, Application US/08475482
GENERAL INFORMATION:
APPLICANT: Chang Yi Wang and Barbara Hosein
TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:

; APPLICATION NUMBER: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 91:
US-09-899-303-91

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACNWTGGERCDLEDRDSEL 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 ACNWTGGERCDLEDRDSEL 20

RESULT 2
US-09-899-303A-91
; Sequence 91, Application US/09899303A
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303A
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 91:
US-09-899-303A-91

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACNWTGGERCDLEDRDSEL 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 ACNWTGGERCDLEDRDSEL 20

RESULT 3
US-09-973-025-91
; Sequence 91, Application US/09973025
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/973,025
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 91:
US-09-973-025-91

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACNWTGGERCDLEDRDSEL 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 ACNWTGGERCDLEDRDSEL 20

RESULT 4
US-09-995-791-91

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 163.1 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-91

Perfect score: 20

Sequence: 1 ACNWTGECDELRDRSEL 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 90918778 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_AA_Main:*

- 1: /cgn2_6/ptodata/1/paa/US06 COMB.pcp.*
- 2: /cgn2_6/ptodata/1/paa/US07 COMB.pcp.*
- 3: /cgn2_6/ptodata/1/paa/US08 COMB.pcp.*
- 4: /cgn2_6/ptodata/1/paa/US09 COMB.pcp.*
- 5: /cgn2_6/ptodata/1/paa/US10 COMB.pcp.*
- 6: /cgn2_6/ptodata/1/paa/US11 COMB.pcp.*
- 7: /cgn2_6/ptodata/1/paa/US12 COMB.pcp.*
- 8: /cgn2_6/ptodata/1/paa/US13 COMB.pcp.*
- 9: /cgn2_6/ptodata/1/paa/US14 COMB.pcp.*
- 10: /cgn2_6/ptodata/1/paa/US15 COMB.pcp.*
- 11: /cgn2_6/ptodata/1/paa/US16 COMB.pcp.*
- 12: /cgn2_6/ptodata/1/paa/US17 COMB.pcp.*
- 13: /cgn2_6/ptodata/1/paa/US18 COMB.pcp.*
- 14: /cgn2_6/ptodata/1/paa/US19 COMB.pcp.*
- 15: /cgn2_6/ptodata/1/paa/US20 COMB.pcp.*
- 16: /cgn2_6/ptodata/1/paa/US21 COMB.pcp.*
- 17: /cgn2_6/ptodata/1/paa/US22 COMB.pcp.*
- 18: /cgn2_6/ptodata/1/paa/US23 COMB.pcp.*
- 19: /cgn2_6/ptodata/1/paa/US24 COMB.pcp.*
- 20: /cgn2_6/ptodata/1/paa/US25 COMB.pcp.*
- 21: /cgn2_6/ptodata/1/paa/US26 COMB.pcp.*
- 22: /cgn2_6/ptodata/1/paa/US27 COMB.pcp.*
- 23: /cgn2_6/ptodata/1/paa/US28 COMB.pcp.*
- 24: /cgn2_6/ptodata/1/paa/US29 COMB.pcp.*
- 25: /cgn2_6/ptodata/1/paa/US30 COMB.pcp.*
- 26: /cgn2_6/ptodata/1/paa/US31 COMB.pcp.*
- 27: /cgn2_6/ptodata/1/paa/US32 COMB.pcp.*
- 28: /cgn2_6/ptodata/1/paa/US33 COMB.pcp.*
- 29: /cgn2_6/ptodata/1/paa/US34 COMB.pcp.*
- 30: /cgn2_6/ptodata/1/paa/US35 COMB.pcp.*
- 31: /cgn2_6/ptodata/1/paa/US36 COMB.pcp.*
- 32: /cgn2_6/ptodata/1/paa/US37 COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	23	US-09-899-303-91

2	20	100.0	20	23	US-09-899-303A-91	Sequence 91, Appl
3	20	100.0	20	25	US-09-973-025-91	Sequence 91, Appl
4	20	100.0	20	25	US-09-995-791-91	Sequence 91, Appl
5	20	100.0	20	25	US-09-995-808-91	Sequence 91, Appl
6	20	100.0	20	25	US-09-995-860-91	Sequence 91, Appl
7	20	100.0	20	26	US-10-020-510-91	Sequence 91, Appl
8	20	100.0	20	29	US-10-321-798-91	Sequence 91, Appl
9	14	70.0	22	9	US-08-507-740-20	Sequence 20, Appl
10	14	70.0	22	13	US-08-905-054-20	Sequence 20, Appl
11	14	70.0	22	13	US-08-905-054B-20	Sequence 20, Appl
12	14	70.0	23	8	US-08-475-482-45	Sequence 45, Appl
13	14	70.0	23	8	US-08-477-072-45	Sequence 45, Appl
14	14	70.0	23	8	US-08-477-582-45	Sequence 45, Appl
15	14	70.0	23	8	US-08-480-253-45	Sequence 45, Appl
16	14	70.0	31	8	US-08-475-482-46	Sequence 46, Appl
17	14	70.0	31	8	US-08-477-072-46	Sequence 46, Appl
18	14	70.0	31	8	US-08-477-582-46	Sequence 46, Appl
19	14	70.0	31	8	US-08-480-253-46	Sequence 46, Appl
20	14	70.0	33	19	US-09-566-266-31	Sequence 31, Appl
21	14	70.0	33	19	US-09-566-266B-31	Sequence 31, Appl
22	14	70.0	33	29	US-10-318-200-18	Sequence 18, Appl
23	14	70.0	39	8	US-08-475-482-47	Sequence 47, Appl
24	14	70.0	39	8	US-08-477-072-47	Sequence 47, Appl
25	14	70.0	39	8	US-08-477-582-47	Sequence 47, Appl
26	14	70.0	39	8	US-08-480-253-47	Sequence 47, Appl
27	14	70.0	41	9	US-08-507-740-21	Sequence 21, Appl
28	14	70.0	41	13	US-08-905-054-21	Sequence 21, Appl
29	14	70.0	41	13	US-08-905-054B-21	Sequence 21, Appl
30	14	70.0	83	9	US-08-537-814A-29	Sequence 29, Appl
31	14	70.0	83	13	US-08-942-483-29	Sequence 29, Appl
32	14	70.0	83	15	US-09-182-186A-29	Sequence 29, Appl
33	14	70.0	141	8	US-08-403-590B-56	Sequence 56, Appl
34	14	70.0	141	8	US-08-444-112-56	Sequence 56, Appl
35	14	70.0	179	8	US-08-444-112-57	Sequence 57, Appl
36	14	70.0	179	8	US-08-444-112-77	Sequence 77, Appl
37	14	70.0	290	19	US-09-564-951-4	Sequence 4, Appl
38	14	70.0	290	19	US-09-564-951B-4	Sequence 4, Appl
39	14	70.0	290	27	US-10-128-578B-3	Sequence 3, Appl
40	14	70.0	290	27	US-10-128-578A-3	Sequence 3, Appl
41	14	70.0	292	19	US-09-564-951A-4	Sequence 4, Appl
42	14	70.0	301	27	US-10-128-578B-5	Sequence 5, Appl
43	14	70.0	301	27	US-10-128-578A-5	Sequence 5, Appl
44	14	70.0				
45	14	70.0				

ALIGNMENTS

RESULT 1

US-09-899-303-91
; Sequence 91, Application US/09899303
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

THIS PAGE BLANK (11/20/20)

```
Query Match      70.0%; Score 14; DB 11; Length 314;
Best Local Similarity 100.0%; Pred. No. 6.9e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 GERCDELDLDRSEL 20
Db      286 GERCDELDLDRSEL 299

RESULT 14
US-09-995-860-42
; Sequence 42, Application US/09995860
; Publication No. US20030118603A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 42
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; US-09-995-860-42

Query Match      70.0%; Score 14; DB 11; Length 314;
Best Local Similarity 100.0%; Pred. No. 6.9e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 GERCDELDLDRSEL 20
Db      286 GERCDELDLDRSEL 299

RESULT 15
US-09-995-791-42
; Sequence 42, Application US/09995791
; Publication No. US20030147918A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 42
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; US-09-995-791-42

Query Match      70.0%; Score 14; DB 12; Length 314;
Best Local Similarity 100.0%; Pred. No. 6.9e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 GERCDELDLDRSEL 20
Db      286 GERCDELDLDRSEL 299
```

Search completed: November 21, 2003, 22:19:41
Job time : 19.5 secs

```

; SEQ ID NO 5
; LENGTH: 301
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-128-590-5

```

```
Query Match      70.0%; Score 14; DB 15; Length 301;
Best Local Similarity 100.0%; Pred. No. 6.7e-07;
Matches 14; Conservative 0; Mismatches 0; Indels
```

Qy 7 GERCLED RDRSEL 20
db 266 GERCLED RDRSEL 279

RESULT 11

US-09-973-025-42
; Sequence 42, Application US/09973025
; Publication No. US20020182706A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEBERT

US-09-973-025-42

```
Query Match      70.0%; Score 14; DB 10; Length 314;
Best Local Similarity 100.0%; Pred. No. 6.9e-07;
Matches 14: Conservative 0; Mismatches 0 Indels
```

Qy 7 GERCLEDRSEL 20
db 286 GERCLEDRSEL 299

RESULT 12

US-09-899-303-42

```

; Sequence 42, Application US/09899303
; Publication No. US20030036110A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
;             BOSMAN, FONS
;             DE MARTYNOFF, GUY
;             BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
;                     PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P. C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 314 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
;
; US-09-899-303-42

```

Query Match 70.0%; Score 14; DB:11; Length 314;
Best Local Similarity 100.0%; Pred.No.6.9e-07;
Matches 14: Conservative 0; Mismatches 0; Indels

Qy 7 GERCLEDRSEL 20
|||
db 286 GERCLEDRSEL 299

RESULT 13

US-09-995-808-42
; Sequence 42, Application US/09995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:

; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-91

Query Match 100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACNWTEGRCLEDRDRSEL 20
|||||
DB 1 ACNWTEGRCLEDRDRSEL 20
|||||

RESULT 6

US-10-318-200-18
; Sequence 18, Application US/10318200
; Publication No. US20030129746A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: DELPA, ERIK
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: EPITOPES IN VIRAL ENVELOPE PROTEINS AND SPECIFIC
; TITLE OF INVENTION: ANTIBODIES DIRECTED AGAINST THESE EPITOPES: USE FOR
; TITLE OF INVENTION: DETECTION OF HCV VIRAL ANTIGEN IN HOST TISSUE
; FILE REFERENCE: 2551-47
; CURRENT APPLICATION NUMBER: US/10/318,200
; CURRENT FILING DATE: 2002-12-13
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US/09/645,470
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: PCT/EP99/02154
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: EP 98870060.5
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 18
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-318-200-18

Query Match 70.0%; Score 14; DB 16; Length 33;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
|||||
DB 9 GERCLEDRDRSEL 22
|||||

RESULT 7

US-10-128-587A-3
; Sequence 3, Application US/10128587A
; Publication No. US20030152940A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Constructs and methods for expression of recombinant HCV envelope
; TITLE OF INVENTION: proteins
; FILE REFERENCE: 134 PCT
; CURRENT APPLICATION NUMBER: US/10/128,587A
; CURRENT FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 290
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-128-587A-3

Query Match 70.0%; Score 14; DB 12; Length 290;
Best Local Similarity 100.0%; Pred. No. 6.5e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
|||||
DB 266 GERCLEDRDRSEL 279
|||||

RESULT 8

US-10-128-590-3
; Sequence 3, Application US/10128590
; Publication No. US20030108561A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Core-glycosylated HCV envelope proteins
; FILE REFERENCE: 135 PCT
; CURRENT APPLICATION NUMBER: US/10/128,590
; CURRENT FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 290
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-128-590-3

Query Match 70.0%; Score 14; DB 15; Length 290;
Best Local Similarity 100.0%; Pred. No. 6.5e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
|||||
DB 266 GERCLEDRDRSEL 279
|||||

RESULT 9

US-10-128-587A-5
; Sequence 5, Application US/10128587A
; Publication No. US20030152940A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Constructs and methods for expression of recombinant HCV envelope
; TITLE OF INVENTION: proteins
; FILE REFERENCE: 134 PCT
; CURRENT APPLICATION NUMBER: US/10/128,587A
; CURRENT FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 301
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-128-587A-5

Query Match 70.0%; Score 14; DB 12; Length 301;
Best Local Similarity 100.0%; Pred. No. 6.7e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
|||||
DB 266 GERCLEDRDRSEL 279
|||||

RESULT 10

US-10-128-590-5
; Sequence 5, Application US/10128590
; Publication No. US20030108561A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Core-glycosylated HCV envelope proteins
; FILE REFERENCE: 135 PCT
; CURRENT APPLICATION NUMBER: US/10/128,590
; CURRENT FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time 19.5 Seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-91

Perfect score: 20

Sequence: 1 ACNWTGRCLEDRSEL 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications_AA.*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	10	US-09-973-025-91
2	20	100.0	20	11	US-09-899-303-91
3	20	100.0	20	11	US-09-995-808-91
4	20	100.0	20	11	US-09-995-860-91
5	20	100.0	20	12	US-09-995-791-91
6	14	70.0	33	16	US-10-318-200-18
7	14	70.0	290	12	US-10-128-587A-3
8	14	70.0	290	15	US-10-128-590-3
9	14	70.0	301	12	US-10-128-587A-5
10	14	70.0	301	15	US-10-128-590-5
11	14	70.0	314	11	US-09-973-025-42
12	14	70.0	314	11	US-09-899-303-42
13	14	70.0	314	11	US-09-995-808-42
14	14	70.0	314	11	US-09-995-860-42
15	14	70.0	314	12	US-09-995-791-42

16	14	70.0	319	10	US-09-973-025-44	Sequence 44, Appl
17	14	70.0	319	11	US-09-899-303-44	Sequence 44, Appl
18	14	70.0	319	11	US-09-995-808-44	Sequence 44, Appl
19	14	70.0	319	11	US-09-995-860-44	Sequence 44, Appl
20	14	70.0	319	12	US-09-995-791-44	Sequence 44, Appl
21	14	70.0	338	10	US-09-973-025-38	Sequence 38, Appl
22	14	70.0	338	11	US-09-899-303-38	Sequence 38, Appl
23	14	70.0	338	11	US-09-995-808-38	Sequence 38, Appl
24	14	70.0	338	11	US-09-995-860-38	Sequence 38, Appl
25	14	70.0	338	12	US-09-995-791-38	Sequence 38, Appl
26	14	70.0	343	10	US-09-973-025-40	Sequence 40, Appl
27	14	70.0	343	11	US-09-899-303-40	Sequence 40, Appl
28	14	70.0	343	11	US-09-995-808-40	Sequence 40, Appl
29	14	70.0	343	11	US-09-995-860-40	Sequence 40, Appl
30	14	70.0	343	12	US-09-995-791-40	Sequence 40, Appl
31	14	70.0	350	10	US-09-923-955-4	Sequence 4, Appl
32	14	70.0	350	14	US-10-104-966-4	Sequence 4, Appl
33	14	70.0	363	10	US-09-407-430-2	Sequence 2, Appl
34	14	70.0	363	12	US-10-128-587A-97	Sequence 97, Appl
35	14	70.0	363	12	US-10-128-587A-98	Sequence 98, Appl
36	14	70.0	363	15	US-10-128-590-97	Sequence 97, Appl
37	14	70.0	363	15	US-10-128-590-98	Sequence 98, Appl
38	14	70.0	463	10	US-09-973-025-46	Sequence 46, Appl
39	14	70.0	463	11	US-09-899-303-46	Sequence 46, Appl
40	14	70.0	463	11	US-09-995-808-46	Sequence 46, Appl
41	14	70.0	463	11	US-09-995-860-46	Sequence 46, Appl
42	14	70.0	463	12	US-09-995-791-46	Sequence 46, Appl
43	14	70.0	490	10	US-09-973-025-36	Sequence 36, Appl
44	14	70.0	490	11	US-09-899-303-36	Sequence 36, Appl
45	14	70.0	490	11	US-09-995-808-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1

US-09-973-025-91
; Sequence 91, Application US/09973025
; Publication No. US20020182706A1

; GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

BOSMAN, FONS

DE MARTYNOFF, GUY

BUYSE, MARIE-ANGEE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHUYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/973,025

FILING DATE: 10-Oct-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

THIS PAGE BLANK (USPTO)

GENERAL INFORMATION:
APPLICANT: Okasinski, Gregory F.
APPLICANT: Schaefer, Verlyn G.
APPLICANT: Suhar, Thomas S.
APPLICANT: Lesniewski, Richard R.
TITLE OF INVENTION: EXPRESSION SYSTEM FOR NON-SECRETOR GENES
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES
STREET: ONE HUNDRED ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,073
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Potembski, Priscilla E.
REGISTRATION NUMBER: 33207
REFERENCE/DOCKET NUMBER: 5763.US.01
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-073-2

Query Match 70.0%; Score 14; DB 3; Length 305;
Best Local Similarity 100.0%; Pred. No. 3.6e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
DB 290 GERCLEDRDRSEL 303

Search completed: November 21, 2003, 21:15:18
Job time : 10.6 secs

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/748,292
FILING DATE: 19910821
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/456,162
FILING DATE: 22-DEC-1989
PRIOR APPLICATION DATA: US 07/610,180
FILING DATE: 07-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: FOREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 4767.US.P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 121 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-748-292-1

Query Match 70.0%; Score 14; DB 1; Length 121;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GERCLEDRDRSEL 20
Db 50 GERCLEDRDRSEL 63

RESULT 13
US-08-444-818-56
Sequence 56 Application US/08444818
Patent No. 6150087
GENERAL INFORMATION:
APPLICANT: Chien, David Y.
APPLICANT: Rutter, William J.
TITLE OF INVENTION: NANBV Diagnostics and Vaccines
NUMBER OF SEQUENCES: 777
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,818
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/403,590
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0110.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508)359-3876
TELEFAX: (508)359-3885

INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 141 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-444-818-56

Query Match 70.0%; Score 14; DB 3; Length 141;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GERCLEDRDRSEL 20
Db 24 GERCLEDRDRSEL 37

RESULT 14
US-08-444-818-77
Sequence 77 Application US/08444818
Patent No. 6150087
GENERAL INFORMATION:
APPLICANT: Chien, David Y.
APPLICANT: Rutter, William J.
TITLE OF INVENTION: NANBV Diagnostics and Vaccines
NUMBER OF SEQUENCES: 777
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,818
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/403,590
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0110.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508)359-3876
TELEFAX: (508)359-3885
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 179 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-444-818-77

Query Match 70.0%; Score 14; DB 3; Length 179;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GERCLEDRDRSEL 20
Db 139 GERCLEDRDRSEL 152

RESULT 15
US-08-478-073-2
Sequence 2 Application US/08478073
Patent No. 6020122

US-07-748-292-6

Query Match 70.0%; Score 14; DB 1; Length 41;
Best Local Similarity 100.0%; Pred. No. 6.1e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
|||||
Db 7 GERCLEDRDRSEL 20

RESULT 10

PCT-US92-07813-21
; Sequence 21, Application PC/TUS9207813
; GENERAL INFORMATION:
; APPLICANT: LESNIEWSKI, RICHARD R.
; APPLICANT: LEUNG, TAT K.
; TITLE OF INVENTION: HEPATITIS C ASSAY
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES CHAD377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: U.S.A.
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/07813
; FILING DATE: 19920916
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKIP, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4767.P3.03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US92-07813-21

Query Match 70.0%; Score 14; DB 5; Length 41;
Best Local Similarity 100.0%; Pred. No. 6.1e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
|||||
Db 7 GERCLEDRDRSEL 20

RESULT 11

US-08-537-811-47
; Sequence 47, Application US/08537811
; Patent No. 5910405
; GENERAL INFORMATION:
; APPLICANT: CHO, JOONG MYUNG
; APPLICANT: CHOI, DEOG YOUNG
; APPLICANT: KIM, CHUN HYUNG
; APPLICANT: SO, HONG SEOB
; APPLICANT: YANG, JAE YOUNG
; APPLICANT: KIM, IN SOO
; APPLICANT: KIM, JOO HO
; TITLE OF INVENTION: IMPROVED HCV DIAGNOSTIC

; TITLE OF INVENTION: AGENTS
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/537,811
; FILING DATE: 24-OCT-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/KR94/00040
; FILING DATE: 29-APR-1994
; APPLICATION NUMBER: KR 93-7440
; FILING DATE: 30-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Jones, III, Harry C
; REGISTRATION NUMBER: 20,280
; REFERENCE/DOCKET NUMBER: 8512-037-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 83 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; OTHER INFORMATION: KHCV E2E, Fig. 2
US-08-537-811-47

Query Match 70.0%; Score 14; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
|||||
Db 3 GERCLEDRDRSEL 16

RESULT 12

US-07-748-292-1
; Sequence 1, Application US/07748292
; Patent No. 5308750
; GENERAL INFORMATION:
; APPLICANT: MEHTA, SMRITI U.
; APPLICANT: JOHNSON, JILL E.
; APPLICANT: DAILEY, STEPHEN H.
; APPLICANT: DESAI, SURESH M.
; APPLICANT: DEVARE, SUSHIL G.
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES TO PUTATIVE HCV
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D-377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

NAME: Maria C. H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4043 US3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 39 amino acids
TYPE: Amino acid
STRANDEDNESS:
TOPOLOGY: Unknown
US-08-262-037-47

Query Match 70.0%; Score 14; DB 1; Length 39;
Best Local Similarity 100.0%; Pred. No. 5.9e-08; Indels 0;
Matches 14; Conservative 0; Mismatches 0; Indels 0;

QY 7 GERCLEDRDRSEL 20
|||||
DB 25 GERCLEDRDRSEL 38
|||||

RESULT 9
US-07-748-292-6
Sequence 6, Application US/07748292
Patent No. 5308750
GENERAL INFORMATION:
APPLICANT: MEHTA, SMRITI U.
APPLICANT: JOHNSON, JILL E.
APPLICANT: DAILEY, STEPHEN H.
APPLICANT: DESAI, SURESH M.
APPLICANT: DEVARE, SUSHIL G.
TITLE OF INVENTION: MONOCLONAL ANTIBODIES TO PUTATIVE HCV
TITLE OF INVENTION: E2/NS1 PROTEINS AND METHODS FOR USING SAME
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES D-377/AP60
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: ILLINOIS
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/748,292
FILING DATE: 19910821
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/456,162
FILING DATE: 22-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/610,180
FILING DATE: 07-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: POREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 4767.US.P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 41 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

NAME: Chang Yi Wang and Barbara Hosein
REGISTRATION NUMBER: 07719,819
REFERENCE/DOCKET NUMBER: 07/667,275
TELECOMMUNICATION INFORMATION:
TELEPHONE: 07/651,735
TELEFAX: 07/558,799
INFORMATION FOR SEQ ID NO: 136
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: AMINO ACID
STRANDEDNESS:

REFERENCE/DOCKET NUMBER: 4767.P3.03
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US92-07813-20

Query Match 70.0%; Score 14; DB 5; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.5e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
|||||
Db 7 GERCLEDRDRSEL 20

RESULT 5

US-08-262-037-45
Sequence 45, Application US/08262037
Patent No. 5747239
GENERAL INFORMATION:
APPLICANT: Chang Yi Wang and Barbara Hosein
TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
NUMBER OF SEQUENCES: 136
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVE.
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/262,037
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/719,819
FILING DATE: 24-June-1991
APPLICATION NUMBER: 07/667,275
FILING DATE: 11-Mar-1991
APPLICATION NUMBER: 07/651,735
FILING DATE: 07-Feb-1991
APPLICATION NUMBER: 07/558,799
FILING DATE: 26-July-1990
APPLICATION NUMBER: 07/510,153
FILING DATE: 16-April-1990
ATTORNEY/AGENT INFORMATION:
NAME: Maria C. H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4043 US3
TELEPHONE: 212-758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: Amino acid
STRANDEDNESS:
TOPOLOGY: Unknown

US-08-262-037-45

Query Match 70.0%; Score 14; DB 1; Length 23;
Best Local Similarity 100.0%; Pred. No. 3.7e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
|||||
Db 9 GERCLEDRDRSEL 22

RESULT 6

US-08-262-037-46
Sequence 46, Application US/08262037
Patent No. 5747239
GENERAL INFORMATION:
APPLICANT: Chang Yi Wang and Barbara Hosein
TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
NUMBER OF SEQUENCES: 136
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVE.
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/262,037
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/719,819
FILING DATE: 24-June-1991
APPLICATION NUMBER: 07/667,275
FILING DATE: 11-Mar-1991
APPLICATION NUMBER: 07/651,735
FILING DATE: 07-Feb-1991
APPLICATION NUMBER: 07/558,799
FILING DATE: 26-July-1990
APPLICATION NUMBER: 07/510,153
FILING DATE: 16-April-1990
ATTORNEY/AGENT INFORMATION:
NAME: Maria C. H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4043 US3
TELEPHONE: 212-758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: Amino acid
STRANDEDNESS:
TOPOLOGY: Unknown

US-08-262-037-46

Query Match 70.0%; Score 14; DB 1; Length 31;
Best Local Similarity 100.0%; Pred. No. 4.8e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
|||||
Db 17 GERCLEDRDRSEL 30

Db 1 ACNWTGGERCDLEDRDRSEL 20

RESULT 2

US-08-927-597-91
; Sequence 91, Application US/08927597
; Patent No. 6245503
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-927-597-91

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.6e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACNWTGGERCDLEDRDRSEL 20
Db 1 ACNWTGGERCDLEDRDRSEL 20

RESULT 3

US-07-748-292-3
; Sequence 3, Application US/07748292
; Patent No. 5308750
; GENERAL INFORMATION:
; APPLICANT: MEHTA, SMRITI U.
; APPLICANT: JOHNSON, JILL E.
; APPLICANT: DAILEY, STEPHEN H.
; APPLICANT: DESAI, SURESH M.
; APPLICANT: DEVARE, SUSHIL G.
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES TO PUTATIVE HCV
; TITLE OF INVENTION: E2/NS1 PROTEINS AND METHODS FOR USING SAME
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ABBOTT LABORATORIES D-377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/748,292
; FILING DATE: 19910821
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/456,162
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/610,180
; FILING DATE: 07-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMBSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4767.US.P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-748-292-3

Query Match 70.0%; Score 14; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.4e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GERCDLEDRDRSEL 20
Db 7 GERCDLEDRDRSEL 20

RESULT 4

PCT-US92-07813-20
; Sequence 20, Application PC/TUS9207813
; GENERAL INFORMATION:
; APPLICANT: LESNIEWSKI, RICHARD R.
; APPLICANT: LEUNG, TAT K.
; TITLE OF INVENTION: HEPATITIS C ASSAY
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES CHAD377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: U.S.A.
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/07813
; FILING DATE: 19920916
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMBSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.6 Seconds
(without alignments)
79.832 Million cell updates/sec

Title: US-09-973-025-91

Perfect score: 20

Sequence: 1 ACNWTGRCLEDRDRSEL 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/1/1aa/PTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-08-612-973-91
2	20	100.0	20	3	US-08-927-597-91
3	14	70.0	21	1	US-07-748-292-3
4	14	70.0	22	5	PCT-US92-07813-20
5	14	70.0	23	1	US-08-262-037-45
6	14	70.0	31	1	US-08-262-037-46
7	14	70.0	33	4	US-09-645-470-18
8	14	70.0	39	1	US-08-262-037-47
9	14	70.0	41	1	US-07-748-292-6
10	14	70.0	41	5	PCT-US92-07813-21
11	14	70.0	83	2	US-08-537-811-47
12	14	70.0	121	1	US-07-748-292-1
13	14	70.0	141	3	US-08-444-818-56
14	14	70.0	179	3	US-08-444-818-77
15	14	70.0	305	3	US-08-478-073-2
16	14	70.0	314	3	US-08-612-973-42
17	14	70.0	314	3	US-08-927-597-42
18	14	70.0	319	3	US-08-612-973-44
19	14	70.0	319	3	US-08-927-597-44
20	14	70.0	333	1	US-08-453-552-12
21	14	70.0	333	1	US-08-710-637-12
22	14	70.0	333	5	PCT-US93-00907-12
23	14	70.0	338	3	US-08-612-973-38
24	14	70.0	338	3	US-08-927-597-38
25	14	70.0	343	3	US-08-612-973-40
26	14	70.0	343	3	US-08-927-597-40
27	14	70.0	367	1	US-08-188-281B-9

28	14	70.0	367	1	US-08-453-552-6	Sequence 6, Appli
29	14	70.0	367	2	US-08-710-637-6	Sequence 6, Appli
30	14	70.0	367	5	PCT-US93-00907-6	Sequence 6, Appli
31	14	70.0	367	5	PCT-US94-07280-9	Sequence 9, Appli
32	14	70.0	367	5	PCT-US95-01087-9	Sequence 9, Appli
33	14	70.0	399	1	US-08-453-552-10	Sequence 10, Appli
34	14	70.0	399	2	US-08-710-637-10	Sequence 10, Appli
35	14	70.0	399	5	PCT-US93-00907-10	Sequence 10, Appli
36	14	70.0	402	1	US-08-460-806-4	Sequence 4, Appli
37	14	70.0	402	1	US-08-460-806-5	Sequence 5, Appli
38	14	70.0	402	1	US-08-460-806-9	Sequence 9, Appli
39	14	70.0	402	1	US-08-460-806-13	Sequence 13, Appli
40	14	70.0	402	1	US-08-460-806-15	Sequence 15, Appli
41	14	70.0	402	1	US-08-460-806-17	Sequence 17, Appli
42	14	70.0	402	1	US-08-325-630-4	Sequence 4, Appli
43	14	70.0	402	1	US-08-325-630-5	Sequence 5, Appli
44	14	70.0	402	1	US-08-325-630-9	Sequence 9, Appli
45	14	70.0	402	1	US-08-325-630-13	Sequence 13, Appli

ALIGNMENTS

RESULT 1
US-08-612-973-91
; Sequence 91, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612.973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-612-973-91

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.6e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACNWTGRCLEDRDRSEL 20
|||||

THIS PAGE BLANK (USPTO)

Query Match 70.0%; Score 14; DB 15; Length 83;
 Best Local Similarity 100.0%; Pred. No. 1.6e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
 |||||
 Db 3 GERCLEDRDRSEL 16

RESULT 14

AAG78134
 ID AAG78134 standard; Protein; 83 AA.

XX AC

XX AAG78134;

XX 04-DEC-2001 (first entry)

DE Hepatitis C virus envelope protein related polypeptide E2E.

XX Hepatitis C virus envelope protein.

XX Unidentified.

XX KR269803-B1.

XX 25-JUL-2000.

XX 17-APR-1993; 93KR-0006494.

XX 17-APR-1993; 93KR-0006494.

XX (GLDS) LG CHEM CO LTD.

XX Cho JM, Kim CH;

XX WPI; 2001-351832/37.

DR N-PSDB; AAH79179.

XX EXPRESSION OF ANTIGEN DETERMINING SITE-FUSION PROTEIN OF HEPATITIS C

PT VIRUS ENVELOPE PROTEIN -

XX Claim 1; Fig 1; 12pp; Korean.

XX The invention relates to expression of an antigen determining site-fusion
 CC protein of the hepatitis C virus envelope protein. The present sequence
 CC is that of a polypeptide of the invention.

XX SQ Sequence 83 AA;

Query Match 70.0%; Score 14; DB 22; Length 83;
 Best Local Similarity 100.0%; Pred. No. 1.6e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
 |||||
 Db 3 GERCLEDRDRSEL 16

RESULT 15

AAR32418

ID AAR32418 standard; peptide; 121 AA.

XX AC

XX AAR32418;

XX 25-MAR-2003 (updated)

DT 03-JUL-1993 (first entry)

XX Sequence of immunogenic domain of E2/NS1 region of hepatitis C virus
 DE (HCV) genome encompassing AAs 600-720.

XX Immunogenic peptide; hepatitis C virus; immunogenic domain;
 KW monoclonal antibody; diagnosis; detection; therapy.

XX

OS Synthetic.

XX WO9304205-A1.

XX 04-MAR-1993.

XX 21-AUG-1992; 92WO-US07189.

XX 21-AUG-1991; 91US-0748292.

XX (ABBO) ABBOTT LAB.

XX Dailey SH, Desai SM, Devare SG, Johnson JE, Mehta SU;

XX WPI; 1993-094030/11.

XX Monoclonal antibodies specific for hepatitis C virus E2-NS1
 PT antigen - useful for diagnosis and evaluation of HCV infections
 PT and in differentiation studies

XX Example; Page 29; 48pp; English.

XX Immunogenic domains of E2/NS1 region of HCV genome encompassing AAs
 CC 600-720 were mapped with PEPSCAN analysis. FAB dimers of IgG
 CC purified from sera of individuals seropositive for antibodies to HCV
 CC proteins were used as the primary antibody for the serological
 CC analysis. Based on the reactivity these sera in EIA, four AA
 CC sequences (AAR32419, R32420, AAR33182, R33183) were identified as
 CC immunogenic domains. Each of these four sequences and an additional
 CC sequence, which was the combination of the two most immunogenic
 CC sequences (AAR33184) were synthesised. Based on the EIA reactivity of
 CC a panel of HCV positive sera, peptide AAR33184 was chosen as the
 CC immunogen for the generation of monoclonal antibodies to HCV NS1.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX SQ Sequence 121 AA;

Query Match 70.0%; Score 14; DB 14; Length 121;
 Best Local Similarity 100.0%; Pred. No. 2.2e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
 |||||
 Db 50 GERCLEDRDRSEL 63

Search completed: November 21, 2003, 20:58:07
 Job time : 32.15 secs

CC acids encoding them, are useful as vaccines to protect humans against
 CC (A) or their mutant strains. The peptides recognise most anti-HCV
 CC antibodies directed against E1 and E2 proteins in the serum; contrast
 CC smaller known peptides or similar peptides expressed in prokaryotes.
 XX
 XX Sequence 33 AA;

Query Match 70.0%; Score 14; DB 20; Length 33;
 Best Local Similarity 100.0%; Pred. No. 7.9e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20
 |||||
 Db 9 GERCLEDRDRSEL 22

RESULT 9

AAR13362
 ID AAR13362 standard; Protein; 41 AA.

XX AAR13362;

AC (ABBO) ABBOTT LAB.

DT 25-MAR-2003 (updated)

DT 23-OCT-1991 (first entry)

XX P643b HCV antigen (643-683).

DE C100-3; hepatitis C virus; immunoassay; epitope.

XX Synthetic.

OS AU9068390-A.

XX AU9068390-A.

PN 27-JUN-1991.

XX 21-DEC-1990; 90AU-0068390.

XX 07-NOV-1990; 90US-0610180.

PR 22-DEC-1989; 89US-0456162.

XX (ABBO) ABBOTT LAB.

PA (LESN/) LESNIEWSKI R R.

XX WPI; 1991-238393/33.

DR Immunological assays for hepatitis C virus antibody - by using

XX polypeptide(s) contg. epitope(s) of hepatitis C virus antigens

PS Claim 1; Page 45; 62pp; English.

XX The polypeptide may be prepared by solid phase synthesis fragment

CC coupling (pref.) or using recombinant technology.

CC The assay has increased sensitivity and is more specific than

CC assays using the polypeptide C100-3 (EP-318216).

CC See also AAQ13146-48 and AAR13343-65.

CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 41 AA;

SQ Query Match 70.0%; Score 14; DB 12; Length 41;

Best Local Similarity 100.0%; Pred. No. 9.4e-07;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCLEDRDRSEL 20

|||||

Db 7 GERCLEDRDRSEL 20

|||||

RESULT 10

AAR33881

ID AAR33881 standard; peptide; 41 AA.

XX AAR33881;

AC

XX 25-MAR-2003 (updated)
 DT 19-JUL-1993 (first entry)
 XX Polypeptide p643b comprising HCV viral antigen.
 XX Hepatitis C virus; NANBH; assay; antibody; p380-JH1; p380-J; p380LG;
 KW p408.
 XX Synthetic.
 XX WO9306247-A1.
 PN 01-APR-1993.
 PD 16-SEP-1992; 92WO-US07813.
 XX 16-SEP-1991; 91US-0760292.
 PR (ABBO) ABBOTT LAB.
 XX Lesniewski RR, Leung TK;
 PI WPI; 1993-117563/14.
 DR Assay for detecting presence of antibody to hepatitis C viral
 PT antigen - by contacting sample with polypeptide contg. at least
 PT one epitope of virus antigen
 XX Disclosure; Page 15; 63pp; English.
 XX The synthetic peptide p643b represents amino acid residues 643-683 of
 CC the hepatitis C viral antigen. The peptide may be used in an assay to
 CC detect antibodies to HCV and thus to diagnose chronic HCV infection.
 CC See also AAR33861-87.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX Sequence 41 AA;
 SQ Query Match 70.0%; Score 14; DB 14; Length 41;
 Best Local Similarity 100.0%; Pred. No. 9.4e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 GERCLEDRDRSEL 20
 |||||
 Db 7 GERCLEDRDRSEL 20
 |||||
 RESULT 11
 AAR33184
 ID AAR33184 standard; peptide; 41 AA.
 XX AAR33184;
 AC 25-MAR-2003 (updated)
 DT 03-JUL-1993 (first entry)
 XX Sequence of peptide 643-683 which is the immunogen for the
 DE generation of monoclonal antibodies to hepatitis
 DE C virus (HCV) NS1.
 XX Immunogenic peptide; hepatitis C virus; immunogenic domain;
 KW monoclonal antibody; diagnosis; detection; therapy.
 XX Synthetic.
 OS WO9304205-A1.
 XX PN 04-MAR-1993.
 XX 21-AUG-1992; 92WO-US07189.
 PF 21-AUG-1991; 91US-0748292.
 PR

PF 25-MAY-2001; 2001WO-AU00622.
 XX
 PR
 XX 26-MAY-2000; 2000AU-0007761.
 XX
 PA (AUSU) UNIV AUSTRALIAN NAT.
 XX
 PI Thomson SA, Ramshaw IA;
 XX
 DR WPI; 2002-147575/19.
 DR N-PSDB; ABK36479.
 XX
 XX New synthetic polypeptides having several different segments of at
 PT least one parent polypeptide linked together differently compared to
 PT the linkage in the parent polypeptide, for inducing immune response
 XX against a pathogen or cancer
 XX
 PS Example 2; Fig 26; 364pp; English.
 XX
 CC The invention relates to a new synthetic polypeptide (I) comprising
 CC several different segments of at least one parent polypeptide linked
 CC together in a different relationship relative to their linkage in the
 CC parent polypeptide to impede, abrogate or otherwise alter at least one
 CC function associated with the parent polypeptide and for inducing an
 CC immune response against a pathogen or cancer. Also included are a
 CC synthetic polynucleotide encoding and a computer system for
 CC designing the synthetic polypeptides. The synthetic polypeptides and
 CC polynucleotides are referred to as a Savine. The synthetic polypeptide is
 CC useful for modulating immune responses preferably directed against a
 CC pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, bone
 CC colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone
 CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.
 CC Compositions comprising the polypeptide may be used in the treatment or
 CC prophylaxis against viral (such as infections caused by HIV (human
 CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
 CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
 CC (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,
 CC Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic
 CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
 CC Trypanosoma, Toxoplasma and Giardia) infections. The present
 CC sequence is a peptide derived from a parent protein used to
 CC construct a savine of the invention.
 XX
 SQ Sequence 30 AA;
 Query Match 70.0%; Score 14; DB 23; Length 30;
 Best Local Similarity 100.0%; Pred. No. 7.3e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 7 GERCLEDLDRSEL 20
 Db 6 GERCLEDLDRSEL 19
 RESULT 7
 AAY39510
 ID AAY39510 standard; peptide; 33 AA.
 AC AAY39510;
 XX
 DT 22-NOV-1999 (first entry)
 XX
 DE HCV E2 protein residues 641-673.
 KW HCV; E1 protein; E2 protein; epitope; hepatitis c virus; antibody;
 XX detection; HCV protein antigen.
 XX
 OS Hepatitis c virus.
 XX
 PN EP947525-A1.
 XX
 PD 06-OCT-1999.
 XX
 PF 27-MAR-1998; 98EP-0870060.
 XX

XX 27-MAR-1998; 98EP-0870060.
 PR (INNO-) INNOGENETICS NV.
 XX
 PA WPI; 1999-542955/46.
 XX
 DR
 XX
 XX New anti-Hepatitis C virus (HCV) antibodies useful for in situ
 PT detection of HCV -
 PT
 XX Example 4; Page 9; 32pp; English.
 PS
 XX This sequence represents residues 641-673 of the hepatitis c virus (HCV)
 CC E2 protein. The invention relates to an antibody (Ab) or its derivative
 CC that specifically binds to the C-terminal region of the HCV E1 protein
 CC (amino acids (aa) 227-383) or the N-terminal region of the HCV E2 protein
 CC (aa 384-450). The Ab allows the in situ detection of HCV protein
 CC antigens. The Ab is used to detect the HCV E1 or E2 proteins in human
 CC peripheral blood cells or liver tissue. It may also be used to detect HCV
 CC in other tissue and fluid samples such as serum, plasma, saliva, mucus
 CC and sections or biopsies such as from skin.
 XX
 SQ Sequence 33 AA;
 Query Match 70.0%; Score 14; DB 20; Length 33;
 Best Local Similarity 100.0%; Pred. No. 7.9e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 7 GERCLEDLDRSEL 20
 Db 9 GERCLEDLDRSEL 22
 RESULT 8
 AAY14189
 ID AAY14189 standard; peptide; 33 AA.
 AC AAY14189;
 XX
 DT 27-JUL-1999 (first entry)
 XX
 DE HCV envelope region peptide C4b.
 XX
 KW HCV; envelope peptide; HCV-related virus; diagnosis; infection; vaccine.
 XX
 OS Hepatitis c virus.
 XX
 PN WO9244466-A2.
 XX
 PD 20-MAY-1999.
 XX
 PF 06-NOV-1998; 98WO-EP07105.
 XX
 PR 06-NOV-1997; 97EP-0870179.
 XX
 PA (INNO-) INNOGENETICS NV.
 XX
 PI Depla E, Maertens G;
 XX
 DR WPI; 1999-327360/27.
 XX
 PT Peptides from the envelope protein of hepatitis C-related viruses
 XX
 PS Claim 2; Page 29; 50pp; English.
 CC This sequence represents a HCV envelope region peptide of the invention.
 CC The peptides: (i) contain more than 20 contiguous amino acids (aa)
 CC derived from the envelope region of a hepatitis C virus (HCV)-related
 CC virus (A); and (ii) bind and recognize anti-(A) antibodies (Ab), also
 CC their functionally equivalent fragments or variants. The peptides are
 CC used: (i) to detect Ab in usual immunoassays, for diagnosing exposure to,
 CC or infection by, (A); (ii) for identifying modulators (II) of peptide-Ab
 CC interaction. Also the peptides, (II) and vectors that express nucleic

CC immunogen for the generation of monoclonal antibodies to HCV NS1.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 21 AA;

Query Match 70.0%; Score 14; DB 14; Length 21;
 Best Local Similarity 100.0%; Pred. No. 5.5e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCDEDRDRSEL 20
 |||||
 Db 7 GERCDEDRDRSEL 20

RESULT 4

AAR1361
 ID AAR13361 standard; Protein; 22 AA.

XX
 AC AAR13361;

XX
 DT 25-MAR-2003 (updated)
 DT 23-OCT-1991 (first entry)

XX
 DE P643a HCV antigen (643-663).

XX
 KW C100-3; hepatitis C virus; immunoassay; epitope.

XX
 OS Synthetic.

PN AU9068390-A.

XX
 PD 27-JUN-1991.

XX
 PF 21-DEC-1990; 90AU-0068390.

XX
 PR 07-NOV-1990; 90US-0610180.

PR 22-DEC-1989; 89US-0456162.

XX
 PA (ABBO) ABBOTT LAB.

PA (LESN/) LESNIEWSKI R R.

XX
 DR WPI; 1991-238393/33.

XX
 PT Immunological assays for hepatitis C virus antibody - by using
 PT polypeptide(s) contg. epitope(s) of hepatitis C virus antigens

XX
 PS Claim 1; Page 45; 62pp; English.

XX
 CC Tyr22 is not part of the HCV sequence but is engineered there for
 CC ease of iodinating the peptide at a later time.
 CC The polypeptide may be prepared by solid phase synthesis fragment
 CC coupling (pref.) or using recombinant technology.
 CC The assay has increased sensitivity and is more specific than
 CC assays using the polypeptide C100-3 (EP-318216).
 CC See also AAQ13146-48 and AAR13343-65.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 XX
 SQ Sequence 22 AA;

Query Match 70.0%; Score 14; DB 12; Length 22;
 Best Local Similarity 100.0%; Pred. No. 5.7e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCDEDRDRSEL 20
 |||||
 Db 7 GERCDEDRDRSEL 20

RESULT 5

AAR33880
 ID AAR33880 standard; peptide; 22 AA.

XX
 AC AAR33880;

XX
 DT 25-MAR-2003 (updated)
 DT 19-JUL-1993 (first entry)
 XX
 DE Polypeptide p643a comprising HCV viral antigen.
 XX
 KW Hepatitis C virus; NANBH; assay; antibody; p380-JH1; p380-J; p380LG;
 KW p408.

XX
 OS Synthetic.

XX
 FH Key Location/Qualifiers
 FT Misc-difference 22

FT /note= "Tyr residue engineered for iodination"

XX
 PN WO9306247-A1.

XX
 PD 01-APR-1993.

XX
 PF 16-SEP-1992; 92WO-US07813.

XX
 PR 16-SEP-1991; 91US-0760292.

XX
 PA (ABBO) ABBOTT LAB.

XX
 PI Lesniewski RR, Leung TK;

XX
 DR WPI; 1993-117563/14.

XX
 PT Assay for detecting presence of antibody to hepatitis C viral
 PT antigen - by contacting sample with polypeptide contg. at least
 PT one epitope of virus antigen

XX
 PS Disclosure; Page 14; 63pp; English.

XX
 CC The synthetic peptide p643a represents amino acid residues 643-663 of
 CC the hepatitis C viral antigen. The peptide may be used in an assay to
 CC detect antibodies to HCV and thus to diagnose chronic HCV infection.
 CC See also AAR33861-87.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 22 AA;

Query Match 70.0%; Score 14; DB 14; Length 22;
 Best Local Similarity 100.0%; Pred. No. 5.7e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GERCDEDRDRSEL 20
 |||||
 Db 7 GERCDEDRDRSEL 20

RESULT 6

AAU84641
 ID AAU84641 standard; Peptide; 30 AA.

XX
 AC AAU84641;

XX
 DT 08-MAY-2002 (first entry)

XX
 DE HCV HepC1a segment 44.

XX
 KW Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
 KW viral infection; human immunodeficiency virus; melanoma;
 KW bacterial infection; Salmonella; Legionella; parasitic infection;
 KW Trypanosoma; Toxoplasma; Giardia.

XX
 OS Hepatitis C virus.

XX
 PN WO200190197-A1.

XX
 PD 29-NOV-2001.

XX

PT assays of HCV

XX Example 7; Page 67; 146pp; English.

PS AAR90965-R90980 and AAR90995-R91015 represent synthetic hepatitis C virus (HCV) E1 and E2 peptides used in competition studies. This sequence represents a synthetic E2 peptide, and corresponds to residues 643-662 of the E2 protein sequence. These sequences are useful for in vitro monitoring of HCV disease, or prognosis of the response to interferon treatment of patients suffering from HCV infection. These sequences compete with the proteins produced by AAT12704-T12709 and AAT12961-T12974, which are included in vectors for the production of recombinant E1, E2, and E1/E2. The recombinant proteins can then be isolated and purified by carrying out a disulphide bond cleavage, or a reduction step with a disulphide bond cleavage agent, after lysis of recombinant host cells. The constructs containing the purified HCV envelope proteins can be used for vaccinating humans against HCV, for in vitro detection of HCV antibodies in a sample, and in a serotyping assay for detecting one or more serological types of HCV present in a biological sample. The constructs can also be immobilised on a solid substrate and incorporated into a reversed phase hybridisation assay for determining the presence or the genotype of HCV. The new purification method preserves the conformation of the recombinantly expressed E1, E2 and E1/E2, and eliminates contaminating proteins. Antigens isolated using this method are more reactive with human sera than those isolated by known techniques.

SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 17; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.9e-13; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACNWTGRCLEDRDRSEL 20
|||||
DB 1 ACNWTGRCLEDRDRSEL 20
|||||

RESULT 2
AAO18720

ID AAO18720 standard; Peptide; 20 AA.

AC AAO18720;

XX 24-OCT-2002 (first entry)

DT Hepatitis C virus E2 protein derived peptide E2-33.

DE Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide;

XX immunostimulant; vaccine.

OS Hepatitis C virus.

XX WO200255548-A2.

PN 18-JUL-2002.

PD 11-JAN-2002; 2002WO-EP00219.

PF 11-JAN-2001; 2001US-260699P.

XX 30-AUG-2001; 2001US-315768P.

XX (INNO-) INNOGENETICS NV.

PA Maertens G, Bosman F, Buyse M;

PI WPI; 2002-599657/64.

DR New therapeutic vaccine compositions comprising at least one purified recombinant hepatitis C virus (HCV) single or specific oligomeric recombinant envelope protein E1 or E2, useful for immunizing humans from HCV infection

PT

XX

PS Example 7; Page 231; 243pp; English.

XX The present invention relates to new therapeutic vaccine compositions for inducing hepatitis C virus (HCV)-specific antibodies, comprising a composition containing at least one purified recombinant HCV single or specific oligomeric recombinant envelope proteins selected from an E1 and an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are useful for inducing HCV-specific antibodies or for immunising humans against HCV. The recombinant HCV E1 and/or E2 proteins are useful as vaccines or therapeutics, in HCV screening and confirmatory antibody tests, for raising antibodies, in the preparation of medicament, and for in vitro monitoring of HCV disease or prognosing the response to treatment of patients suffering from HCV infection. The present sequence is a peptide derived from the proteins of the invention.

SQ Sequence 20 AA;

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.9e-13; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACNWTGRCLEDRDRSEL 20
|||||
DB 1 ACNWTGRCLEDRDRSEL 20
|||||

RESULT 3
AAR32420

ID AAR32420 standard; peptide; 21 AA.

XX AAR32420;

AC 25-MAR-2003 (updated)

XX 03-JUL-1993 (first entry)

DT Sequence of immunogenic domain of E2/NS1 region of hepatitis C virus (HCV) genome encompassing AAs 643-663.

DE Immunogenic peptide; hepatitis C virus; immunogenic domain;

XX monoclonal antibody; diagnosis; detection; therapy.

OS Synthetic.

XX WO9304205-A1.

PN 04-MAR-1993.

XX 21-AUG-1992; 92WO-US07189.

XX 21-AUG-1991; 91US-0748292.

XX (ABBO) ABBOTT LAB.

XX Dailey SH, Desai SM, Devare SG, Johnson JE, Mehta SU;

PI WPI; 1993-094030/11.

DR Monoclonal antibodies specific for hepatitis C virus E2-NS1 antigen - useful for diagnosis and evaluation of HCV infections and in differentiation studies

XX Example; Page 30; 48pp; English.

XX Immunogenic domains of E2/NS1 region of HCV genome encompassing AAs 600-720 were mapped with PEPSAN analysis. FAB dimers of Igg purified from sera of individuals seropositive for antibodies to HCV proteins were used as the primary antibody for the serological analysis. Based on the reactivity these sera in EIA, four AA sequences (AAR32419, R32420, AAR33182, R33183) were identified as immunogenic domains. Each of these four sequences and an additional sequence, which was the combination of the two most immunogenic sequences (AAR33184) were synthesised. Based on the EIA reactivity of a panel of HCV positive sera, peptide AAR33184 was chosen as the

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:22 ; Search time 32.1 Seconds
(without alignments)
98.895 Million cell updates/sec

Title: US-09-973-025-91

Perfect score: 20

Sequence: 1 ACNWTPEERCDLEDRRSEL 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	17 AAR91014	HCV E2 peptide E2-
2	20	100.0	20	23 AAO18720	Hepatitis C virus
3	14	70.0	21	14 AAR32420	Sequence of immuno
4	14	70.0	22	12 AAR13361	P643a HCV antigen
5	14	70.0	22	14 AAR33880	Polypeptide p643a
6	14	70.0	30	23 AAU84641	HCV HepC1a segment
7	14	70.0	33	20 AAY39510	HCV E2 protein res
8	14	70.0	33	20 AAT14189	HCV envelope regio
9	14	70.0	41	12 AAR13362	P643b HCV antigen

10	14	70.0	41	14 AAR33881	Polypeptide p643b
11	14	70.0	41	14 AAR33184	Sequence of peptid
12	14	70.0	83	15 AAR62867	Korean-type Hepati
13	14	70.0	83	15 AAR62514	HCV E2E envelope
14	14	70.0	83	22 AAG78134	Hepatitis C virus
15	14	70.0	121	14 AAR32418	Sequence of immuno
16	14	70.0	141	10 AAP90159	Sequence of hepati
17	14	70.0	141	10 AAP92042	Sequence encoded i
18	14	70.0	141	21 AAB18525	Protein encoded by
19	14	70.0	179	21 AAB18529	HCV polypeptide 3..
20	14	70.0	201	13 AAR25856	Protein encoded by
21	14	70.0	203	19 AAW68464	HCV NS1(gp70) pept
22	14	70.0	209	13 AAR29553	HCV NS1(gp70) pept
23	14	70.0	209	13 AAR29554	HCV NS1(gp70) pept
24	14	70.0	209	13 AAR29555	HCV NS1(gp70) pept
25	14	70.0	290	24 AAE32843	Hepatitis C virus
26	14	70.0	290	24 AAO26541	HCV envelope prote
27	14	70.0	290	24 ABP55530	Hepatitis C virus
28	14	70.0	300	21 ABB08497	Envelope 2 protein
29	14	70.0	301	24 AAE32845	Hepatitis C virus
30	14	70.0	301	24 AAO26543	HCV envelope prote
31	14	70.0	301	24 ABP55532	HCV type 1b Els pr
32	14	70.0	304	21 AAY69651	HCV mature HVRI-de
33	14	70.0	305	18 AAW00929	Recombinant HCV E2
34	14	70.0	307	21 AAY69650	HCV mature HVRI-de
35	14	70.0	314	23 AAO18675	Hepatitis C virus
36	14	70.0	319	23 AAO18676	Hepatitis C virus
37	14	70.0	325	13 AAR29557	gp70 peptide N19MX
38	14	70.0	325	13 AAR29558	gp70 peptide N19MX
39	14	70.0	333	14 AAR40118	HGH-HCV-E2 fusion
40	14	70.0	338	23 AAO18673	Hepatitis C virus
41	14	70.0	343	23 AAO18674	Hepatitis C virus
42	14	70.0	350	23 AAE19891	Hepatitis C virus
43	14	70.0	363	22 AAB68042	Amino acid sequenc
44	14	70.0	363	24 AAE32880	Hepatitis C virus
45	14	70.0	363	24 AAE32881	Hepatitis C virus

ALIGNMENTS

RESULT 1
AAR91014
ID AAR91014 standard; peptide; 20 AA.
XX
AC AAR91014;
XX
DT 25-SEP-1996 (first entry)
XX
DE HCV E2 peptide E2-33 for competition studies.
XX
DE HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;
KW serotype; reversed phase hybridisation assay; genotype; antigen; sera.
XX
OS Synthetic.
XX
PN WO9604385-A2.
XX
PD 15-FEB-1996.
XX
PF 31-JUL-1995; 95WO-EP03031.
XX
PR 29-JUL-1994; 94EP-0870132.
XX
PA (INNO-) INNOGENETICS NV.
XX
PI Bosman F, Buysse M, De Martynoff G, Maertens G;
XX
DR WPI; 1996-129401/13.
XX
PT Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope
PT proteins - in presence of di: sulphide bond cleavage agent, to
PT produce proteins suitable for direct use in vaccines or diagnostic

THIS PAGE BLANK (USPTO)

KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1 1
 FT NON_TER 74 74
 SQ SEQUENCE 74 AA; 8454 MW; 6C251D9B130CFC26 CRC64;
 Query Match 85.0%; Score 17; DB 12; Length 74;
 Best Local Similarity 100.0%; Pred. No. 1.5e-10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MYVGGVEHRLEAACNWT 17
 Db 3 MYVGGVEHRLEAACNWT 19
 RESULT 13
 Q9E3G1 PRELIMINARY; PRT; 74 AA.
 AC Q9E3G1;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE E2 glycoprotein-PKR-eif-2 (genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2.9;
 RX MEDLINE=20375028; PubMed=10915068;
 RA Polyak S.J., Nausbaum J.B., Larson A.M., Cotler S., Carithers Jr R.L.,
 RA Gretch D.R.;
 RT "The protein kinase-interacting domain in the hepatitis C virus
 RT envelope glycoprotein-2 gene is highly conserved in genotype 1-
 RT infected patients treated with interferon."
 RL J. Infect. Dis. 182:397-404(2000).
 DR EMBL; AF264977; AAG17316.1; -.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1 1
 FT NON_TER 74 74
 SQ SEQUENCE 74 AA; 8454 MW; 6C251D9B130CFC26 CRC64;
 Query Match 85.0%; Score 17; DB 12; Length 74;
 Best Local Similarity 100.0%; Pred. No. 1.5e-10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MYVGGVEHRLEAACNWT 17
 Db 3 MYVGGVEHRLEAACNWT 19
 RESULT 14
 Q9E3H1 PRELIMINARY; PRT; 74 AA.
 AC Q9E3H1;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE E2 glycoprotein-PKR-eif-2 (genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=11Dre;
 RX MEDLINE=20375028; PubMed=10915068;
 RA Polyak S.J., Nausbaum J.B., Larson A.M., Cotler S., Carithers Jr R.L.,

RA Gretch D.R.;
 RT "The protein kinase-interacting domain in the hepatitis C virus
 RT envelope glycoprotein-2 gene is highly conserved in genotype 1-
 RT infected patients treated with interferon."
 RL J. Infect. Dis. 182:397-404(2000).
 DR EMBL; AF264967; AAG17306.1; -.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1 1
 FT NON_TER 74 74
 SQ SEQUENCE 74 AA; 8469 MW; 01E4DD9B0DFD1333 CRC64;
 Query Match 85.0%; Score 17; DB 12; Length 74;
 Best Local Similarity 100.0%; Pred. No. 1.5e-10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MYVGGVEHRLEAACNWT 17
 Db 3 MYVGGVEHRLEAACNWT 19
 RESULT 15
 Q9E3G8 PRELIMINARY; PRT; 74 AA.
 AC Q9E3G8;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE E2 glycoprotein-PKR-eif-2 (genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=11-29;
 RX MEDLINE=20375028; PubMed=10915068;
 RA Polyak S.J., Nausbaum J.B., Larson A.M., Cotler S., Carithers Jr R.L.,
 RA Gretch D.R.;
 RT "The protein kinase-interacting domain in the hepatitis C virus
 RT envelope glycoprotein-2 gene is highly conserved in genotype 1-
 RT infected patients treated with interferon."
 RL J. Infect. Dis. 182:397-404(2000).
 DR EMBL; AF264970; AAG17309.1; -.
 DR InterPro; IPR002531; HCV NS1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1 1
 FT NON_TER 74 74
 SQ SEQUENCE 74 AA; 8453 MW; 11A4DD9B0DFD1333 CRC64;
 Query Match 85.0%; Score 17; DB 12; Length 74;
 Best Local Similarity 100.0%; Pred. No. 1.5e-10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MYVGGVEHRLEAACNWT 17
 Db 3 MYVGGVEHRLEAACNWT 19

Search completed: November 21, 2003, 21:08:22
 Job time : 25.45 secs

```

AC Q9E3G6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2 glycoprotein-PKR-eIF-2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=11-27;
RX MEDLINE=20375028; PubMed=10915068;
RA Polyak S.J., Nausbaum J.B., Larson A.M., Cotler S., Carithers Jr R.L.,
RA 'Gretch D.R.;
RT "The protein kinase-interacting domain in the hepatitis C virus
RT envelope glycoprotein-2 gene is highly conserved in genotype 1-
RT infected patients treated with interferon.";
RL J. Infect. Dis. 182:397-404(2000).
DR EMBL; AF264972; AAG17311.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI.
DR ProDom; PD186062; HCV_NSI.
DR Coats protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 74
FT NON_TER 74
SQ SEQUENCE 74 AA; 8469 MW; 01E4DD9B0DFD1333 CRC64;

Query Match 85.0%; Score 17; DB 12; Length 74;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYVGGVEHRLAACNWT 17
Db 3 MYVGGVEHRLAACNWT 19

RESULT 10
Q9E3G9 PRELIMINARY; PRT; 74 AA.
ID Q9E3G9;
AC Q9E3G9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2 glycoprotein-PKR-eIF-2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=19r6;
RX MEDLINE=20375028; PubMed=10915068;
RA Polyak S.J., Nausbaum J.B., Larson A.M., Cotler S., Carithers Jr R.L.,
RA 'Gretch D.R.;
RT "The protein kinase-interacting domain in the hepatitis C virus
RT envelope glycoprotein-2 gene is highly conserved in genotype 1-
RT infected patients treated with interferon.";
RL J. Infect. Dis. 182:397-404(2000).
DR EMBL; AF264969; AAG17308.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI.
DR ProDom; PD186062; HCV_NSI.
DR Coats protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 74
FT NON_TER 74
SQ SEQUENCE 74 AA; 8440 MW; 199E71EB130CFC27 CRC64;

Query Match 85.0%; Score 17; DB 12; Length 74;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 MYVGGVEHRLAACNWT 17
Db 3 MYVGGVEHRLAACNWT 19

RESULT 11
Q9E3F5 PRELIMINARY; PRT; 74 AA.
ID Q9E3F5;
AC Q9E3F5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2 glycoprotein-PKR-eIF-2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2-24;
RX MEDLINE=20375028; PubMed=10915068;
RA Polyak S.J., Nausbaum J.B., Larson A.M., Cotler S., Carithers Jr R.L.,
RA 'Gretch D.R.;
RT "The protein kinase-interacting domain in the hepatitis C virus
RT envelope glycoprotein-2 gene is highly conserved in genotype 1-
RT infected patients treated with interferon.";
RL J. Infect. Dis. 182:397-404(2000).
DR EMBL; AF264983; AAG17322.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI.
DR ProDom; PD186062; HCV_NSI.
DR Coats protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 74
FT NON_TER 74
SQ SEQUENCE 74 AA; 8438 MW; 7C651D9B130CFC26 CRC64;

```

```

Query Match 85.0%; Score 17; DB 12; Length 74;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYVGGVEHRLAACNWT 17
Db 3 MYVGGVEHRLAACNWT 19

RESULT 12
Q9E3F7 PRELIMINARY; PRT; 74 AA.
ID Q9E3F7;
AC Q9E3F7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2 glycoprotein-PKR-eIF-2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2-3;
RX MEDLINE=20375028; PubMed=10915068;
RA Polyak S.J., Nausbaum J.B., Larson A.M., Cotler S., Carithers Jr R.L.,
RA 'Gretch D.R.;
RT "The protein kinase-interacting domain in the hepatitis C virus
RT envelope glycoprotein-2 gene is highly conserved in genotype 1-
RT infected patients treated with interferon.";
RL J. Infect. Dis. 182:397-404(2000).
DR EMBL; AF264981; AAG17320.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI.
DR ProDom; PD186062; HCV_NSI.

```

```

RL J. Infect. Dis. 182:397-404(2000).
DR EMBL; AF264979; AAG17318.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 74
SQ SEQUENCE 74 AA; 8438 MW; 7C651D9B130CFC26 CRC64;

Query Match 85.0%; Score 17; DB 12; Length 74;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLEAACNWT 17
Db 3 MYVGGVEHRLEAACNWT 19

RESULT 6
Q9E3G5 PRELIMINARY; PRT; 74 AA.
AC Q9E3G5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2 glycoprotein-PKR-eIF-2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=11-26;
RX MEDLINE=20375028; PubMed=10915068;
RA Polyak S.J.; Nausbaum J.B.; Larson A.M.; Cotler S.; Carithers Jr R.L.;
RA Gretch D.R.;
RT "The protein kinase-interacting domain in the hepatitis C virus
RT envelope glycoprotein-2 gene is highly conserved in genotype 1-
RT infected patients treated with interferon."
RL J. Infect. Dis. 182:397-404(2000).
DR EMBL; AF264973; AAG17312.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 74
SQ SEQUENCE 74 AA; 8453 MW; 11A4DD9B0DFD1333 CRC64;

Query Match 85.0%; Score 17; DB 12; Length 74;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLEAACNWT 17
Db 3 MYVGGVEHRLEAACNWT 19

RESULT 7
Q9E3H8 PRELIMINARY; PRT; 74 AA.
AC Q9E3H8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2 glycoprotein-PKR-eIF-2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=11-26;
RX MEDLINE=20375028; PubMed=10915068;
RA Polyak S.J.; Nausbaum J.B.; Larson A.M.; Cotler S.; Carithers Jr R.L.;
RA Gretch D.R.;
RT "The protein kinase-interacting domain in the hepatitis C virus
RT envelope glycoprotein-2 gene is highly conserved in genotype 1-
RT infected patients treated with interferon."
RL J. Infect. Dis. 182:397-404(2000).
DR EMBL; AF264973; AAG17312.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 74
SQ SEQUENCE 74 AA; 8453 MW; 11A4DD9B0DFD1333 CRC64;

Query Match 85.0%; Score 17; DB 12; Length 74;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLEAACNWT 17
Db 3 MYVGGVEHRLEAACNWT 19

RESULT 8
Q9E3G3 PRELIMINARY; PRT; 74 AA.
AC Q9E3G3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2 glycoprotein-PKR-eIF-2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=11-18;
RX MEDLINE=20375028; PubMed=10915068;
RA Polyak S.J.; Nausbaum J.B.; Larson A.M.; Cotler S.; Carithers Jr R.L.;
RA Gretch D.R.;
RT "The protein kinase-interacting domain in the hepatitis C virus
RT envelope glycoprotein-2 gene is highly conserved in genotype 1-
RT infected patients treated with interferon."
RL J. Infect. Dis. 182:397-404(2000).
DR EMBL; AF264975; AAG17314.1; -.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; 1.
DR ProDom; PD186062; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 74
SQ SEQUENCE 74 AA; 8422 MW; 1014DD9B0DFD1323 CRC64;

Query Match 85.0%; Score 17; DB 12; Length 74;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLEAACNWT 17
Db 3 MYVGGVEHRLEAACNWT 19

RESULT 9
Q9E3G6 PRELIMINARY; PRT; 74 AA.
ID Q9E3G6

```

```

SQ SEQUENCE 74 AA; 8438 MW; 7C651D9B130CFC26 CRC64;
Query Match 85.0%; Score 17; DB 12; Length 74;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLEAACNWT 17
DB 3 MYVGGVEHRLEAACNWT 19

RESULT 4
Q9E3H7 PRELIMINARY; PRT; 74 AA.
AC Q9E3H7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2 glycoprotein-PKR-eif-2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4m05;
RX MEDLINE=20375028; PubMed=10915068;
RA Polyak S.J., Nausbaum J.B., Larson A.M., Cotler S., Carithers Jr R.L.,
RA Gretch D.R.;
RT "The protein kinase-interacting domain in the hepatitis C virus
RT envelope glycoprotein-2 gene is highly conserved in genotype 1-
RT infected patients treated with interferon.";
RL J. Infect. Dis. 182:397-404(2000).
DR EMBL; AF264966; AACG17305.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 74
SQ SEQUENCE 74 AA; 8469 MW; 01E4DD9B0DFD1333 CRC64;

Query Match 85.0%; Score 17; DB 12; Length 74;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLEAACNWT 17
DB 3 MYVGGVEHRLEAACNWT 19

RESULT 3
Q9E3G0 PRELIMINARY; PRT; 74 AA.
AC Q9E3G0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2 glycoprotein-PKR-eif-2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2-8;
RX MEDLINE=20375028; PubMed=10915068;
RA Polyak S.J., Nausbaum J.B., Larson A.M., Cotler S., Carithers Jr R.L.,
RA Gretch D.R.;
RT "The protein kinase-interacting domain in the hepatitis C virus
RT envelope glycoprotein-2 gene is highly conserved in genotype 1-
RT infected patients treated with interferon.";
RL J. Infect. Dis. 182:397-404(2000).
DR EMBL; AF264978; AACG17317.1; -.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 74

Query Match 85.0%; Score 17; DB 12; Length 74;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLEAACNWT 17
DB 3 MYVGGVEHRLEAACNWT 19

RESULT 5
Q9E3F9 PRELIMINARY; PRT; 74 AA.
AC Q9E3F9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2 glycoprotein-PKR-eif-2 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2-7;
RX MEDLINE=20375028; PubMed=10915068;
RA Polyak S.J., Nausbaum J.B., Larson A.M., Cotler S., Carithers Jr R.L.,
RA Gretch D.R.;
RT "The protein kinase-interacting domain in the hepatitis C virus
RT envelope glycoprotein-2 gene is highly conserved in genotype 1-
RT infected patients treated with interferon.";

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:45:57 ; Search time 24.4 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-90

Perfect score: 20

Sequence: 1 MYVGVGVEHRLAECACNWTPE 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_prodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	85.0	73	12 Q9E3I1	Q9E3I1 hepatitis c
2	17	85.0	74	12 Q9E3H2	Q9E3H2 hepatitis c
3	17	85.0	74	12 Q9E3G0	Q9E3G0 hepatitis c
4	17	85.0	74	12 Q9E3H7	Q9E3H7 hepatitis c
5	17	85.0	74	12 Q9E3F9	Q9E3F9 hepatitis c
6	17	85.0	74	12 Q9E3G5	Q9E3G5 hepatitis c
7	17	85.0	74	12 Q9E3H8	Q9E3H8 hepatitis c
8	17	85.0	74	12 Q9E3G3	Q9E3G3 hepatitis c
9	17	85.0	74	12 Q9E3G6	Q9E3G6 hepatitis c
10	17	85.0	74	12 Q9E3G9	Q9E3G9 hepatitis c
11	17	85.0	74	12 Q9E3F5	Q9E3F5 hepatitis c
12	17	85.0	74	12 Q9E3F7	Q9E3F7 hepatitis c
13	17	85.0	74	12 Q9E3G1	Q9E3G1 hepatitis c
14	17	85.0	74	12 Q9E3H1	Q9E3H1 hepatitis c
15	17	85.0	74	12 Q9E3G8	Q9E3G8 hepatitis c
16	17	85.0	74	12 Q9E3H9	Q9E3H9 hepatitis c

17	17	85.0	74	12	Q9E3G7	Q9E3G7 hepatitis c
18	17	85.0	74	12	Q9E3F8	Q9E3F8 hepatitis c
19	17	85.0	74	12	Q9E3I4	Q9E3I4 hepatitis c
20	17	85.0	74	12	Q9E3F6	Q9E3F6 hepatitis c
21	17	85.0	74	12	Q9E3I0	Q9E3I0 hepatitis c
22	17	85.0	74	12	Q9E3G4	Q9E3G4 hepatitis c
23	17	85.0	74	12	Q9E3H5	Q9E3H5 hepatitis c
24	17	85.0	74	12	Q9E3G2	Q9E3G2 hepatitis c
25	17	85.0	74	12	Q9E3H4	Q9E3H4 hepatitis c
26	17	85.0	129	12	Q8JYQ1	Q8JYQ1 hepatitis c
27	17	85.0	129	12	Q9IEE0	Q9IEE0 hepatitis c
28	17	85.0	129	12	Q9IEG8	Q9IEG8 hepatitis c
29	17	85.0	129	12	Q9IEG9	Q9IEG9 hepatitis c
30	17	85.0	129	12	Q9IED9	Q9IED9 hepatitis c
31	17	85.0	206	12	Q8I571	Q8I571 hepatitis c
32	17	85.0	328	12	Q9WHJ5	Q9WHJ5 hepatitis c
33	17	85.0	420	12	Q98UN4	Q98UN4 hepatitis c
34	17	85.0	452	12	Q9QD17	Q9QD17 hepatitis c
35	17	85.0	778	12	Q04184	Q04184 hepatitis c
36	17	85.0	783	12	Q68952	Q68952 hepatitis c
37	17	85.0	2436	12	Q81756	Q81756 hepatitis c
38	17	85.0	3010	12	Q68788	Q68788 hepatitis c
39	17	85.0	3010	12	Q9J3H7	Q9J3H7 hepatitis c
40	17	85.0	3010	12	Q9J3H9	Q9J3H9 hepatitis c
41	17	85.0	3010	12	Q9QP06	Q9QP06 hepatitis c
42	17	85.0	3010	12	Q81757	Q81757 hepatitis c
43	17	85.0	3010	12	Q9WMX2	Q9WMX2 hepatitis c
44	17	85.0	3011	12	Q9IFES	Q9IFES hepatitis c
45	17	85.0	3011	12	Q36579	Q36579 hepatitis c

ALIGNMENTS

RESULT 1

ID	Q9E3I1	PRELIMINARY;	PRT;	73 AA.
AC	Q9E3I1;			
DT	01-MAR-2001 (TREMELrel. 16, Created)			
DT	01-MAR-2001 (TREMELrel. 16, Last sequence update)			
DE	01-Oct-2002 (TREMELrel. 22, Last annotation update)			
DE	E2 glycoprotein-EPR-eIF-2 (Genome polyprotein) (Fragment).			
OS	Hepatitis C virus.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC	Hepacivirus.			
OX	NCBI_TaxID=11103;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=5wk6;			
RX	MEDLINE=20375028; PubMed=10915068;			
RA	Polyak S.J., Noursbaum J.B., Larson A.M., Colter S., Carithers Jr R.L.,			
RA	Gretch D.R.;			
RT	"The protein kinase-interacting domain in the hepatitis C virus			
RT	envelope glycoprotein-2 gene is highly conserved in genotype 1-			
RT	infected patients treated with interferon."			
RL	J. Infect. Dis. 182:397-404(2000).			
DR	EMBL: AF264957; AAG17296.1;			
DR	InterPro; IPR002531; HCV_NSI.			
DR	Pfam; PF01560; HCV_NSI; 1.			
DR	ProDom; PD186062; HCV_NSI; 1.			
KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;			
KW	Polyprotein; Transmembrane.			
FT	NON_TER 1			
FT	NON_TER 73			
SQ	SEQUENCE 73 AA; 8340 MW; 251D9B130CFC260C CRC64;			
Query Match 85.0%; Score 17; DB 12; Length 73;				
Best Local Similarity 100.0%; Pred. No. 1.5e-10;				
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1 MYVGVGVEHRLAECACNWT 17			
DB	3 MYVGVGVEHRLAECACNWT 19			

THIS PAGE BLANK (USPTO)

Search completed: November 21, 2003, 21:00:03
Job time : 5.85 secs

KW	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW	Complete proteome.
FT	SITE 70
FT	SITE 81
FT	SITE 727
FT	SITE 731
FT	BINDING 730
FT	ATP (BY SIMILARITY) .
SQ	SEQUENCE 957 AA; 107399 MW; AGFFBF26AAPBC1FF CRC64;
	"HIGH" REGION.
	"KMSKS" REGION.
	ATP (BY SIMILARITY) .
	Query Match 35.0%; Score 7; DB i; Length 957;
	Best Local Similarity 100.0%; Pred.No. 4.5;
	Matches 7; Conservative 0; Mismatches 0;
	Indels 0; Gaps 0;

```

Query Match          30.0%; Score 6; DB 1; Length 107;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps
0;

QY      9 RLEAAC 14
        |||||
Db      69 RLEAAC 74

RESULT 15
ID_ MAIZE          STANDARD;      PRT;      168 AA.
DH1_ MAIZE
AC      P12950; P16318;
DT      01-JAN-1990 (Rel. 13, Created)
DT      01-AUG-1990 (Rel. 15, Last sequence update)
DT      01-JUN-1994 (Rel. 29, Last annotation update)
DE      Dehydrin DHN1 (M3) (RAB-17 protein).

```

RESULT	15
DH1_MAIZE	
ID_DH1_MAIZE	STANDARD; PRT; 168 AA.
AC	P12950; P16318;
DT	01-JAN-1990 (Rel. 13, Crested)
DT	01-AUG-1990 (Rel. 15, Last sequence update)
DT	01-JUN-1994 (Rel. 29, Last annotation update)
DE	Dehydrin DHN1 (M3) (RAB-17 protein).
GN	DHN1.
OS	Zea mays (Maize).
OC	Eukaryota: Viridiplantae: Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OC NCBI_TaxID=4577;
RN [1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP MEDLINE=93357436; PubMed=2562763;
RX Close T.J., Kott A.A., Chandler P.M.;
RT "A cDNA-based comparison of denaturation-induced proteins (dehydrins)
RT in barley and corn."
RN Plant Mol. Biol. 13:95-108(1989).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=CV. AC 1503;
RX MEDLINE=91346629; PubMed=2151715;
RX Villardelli J., Goday A., Freire M.A., Torrent M., Martinez M.C.,
RA Torne J.M., Pages M.;
RT "Gene sequence, developmental expression, and protein phosphorylation
RT of RAB-17 in maize."
RN Plant Mol. Biol. 14:423-432(1990).
RN [3]
RN PHOSPHORYLATION BY CKII.
RX MEDLINE=92042196; PubMed=1939268;
RA Plana M., Iartre E., Ertija R., Goday A., Pages M., Martinez M.C.;
RT "Phosphorylation of maize RAB-17 protein by casein kinase 2."
RN J. Biol. Chem. 266:122510-122514(1991).
CC -!- DEVELOPMENTAL STAGE: ABA TREATMENT INDUCED THE SYNTHESIS OF RAB-17
CC IN CALLI, HOWEVER, THE RAB-17 PROTEINS WERE FOUND TO BE HIGHLY
CC PHOSPHORYLATED ONLY IN EMBRYOS.
CC -!- INDUCTION: By abscisic acid (ABA) and water stress.
CC -!- PTM: SERINE IS THE ONLY AMINO ACID WHICH BECOMES PHOSPHORYLATED.
CC -!- SIMILARITY: BELONGS TO THE PLANT DEHYDRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

Qy 3 VGGVEH 8
Dp 24 VGGVEH 29

```

RESULT 11
T408_BURCE STANDARD; PRT; 518 AA.
AC Q51649;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative transposase for insertion sequence IS408.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 17616 / 249;
RX MEDLINE=94302134; PubMed=7518087;
RA Byrne A.M., Lessie T.G.;
RT "Characteristics of IS401, a new member of the IS3 family implicated
RT in plasmid rearrangements in Pseudomonas cepacia.";
RL Plasmid 31:138-147(1994).
CC -!- FUNCTION: REQUIRED FOR THE TRANSPOSITION OF THE INSERTION ELEMENT
CC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE IS21/IS408/IS1162 FAMILY OF
CC TRANSPOSASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP005283; BAC00412.1; --
CC HAMAP; MF_00049; --; 1.
CC InterPro; IPR002302; Leu-trnASyntia.
CC InterPro; IPR002300; trNA-synt_1a.
CC InterPro; IPR001412; trNA-synt_1.
CC Pfam; PF00133; trNA-synt_1; 1.
CC PRINTS; PR00985; TRNASYNTHLEU.
CC TIGRFAMs; TIGR00396; leus_bact; 1.
CC PROSITE; PS00178; AA_TRNA_LIGASE_1; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 66 77 "HIGH" REGION.
FT SITE 722 726 "KMSKS" REGION.
FT BINDING 725 725 ATP (BY SIMILARITY).
FT SEQUENCE 952 AA; 106778 MW; 3A1D2E2CD12D13E5 CRC64;
SQ
Query Match 35.0%; Score 7; DB 1; Length 952;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 YVGVGEH 8
DB 642 YVGVGEH 648
RESULT 13
SYL_COREF STANDARD; PRT; 957 AA.
AC Q8FLM0;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR CE2848.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Usuda Y., Sugimoto S.;
RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP005223; BAC19658.1; --
CC HAMAP; MF_00049; --; 1.
CC Pfam; PF00133; trNA-synt_1; 1.
CC TIGRFAMs; TIGR00396; leus_bact; 1.
CC PROSITE; PS00178; AA_TRNA_LIGASE_1; 1.

```

FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;

Query Match 50.0%; Score 10; DB 1; Length 3033;
Best Local Similarity 100.0%; Pred. No. 0.0079;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYVGGVEHRL 10
Db 635 MYVGGVEHRL 644

RESULT 10
POLG_HCVJ8 STANDARD; PRT; 3033 AA.
ID P26661;
AC P26661;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polypeptide [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.-); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J8) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
ON NCBI_TaxID=11115;
RX [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
RA Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 198;331-341(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polypeptide, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1',
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA) (N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/ebis/sib.ch/announcement/>
CC or send an email to license@ebi.ac.uk).

DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV NS5a.
DR InterPro; IPR007095; RNA pol. PS.
DR InterPro; IPR007094; RNA pol. PSvir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV env; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00998; Viral RdRp; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT TRANSMEM 347 369
FT ACT_SITE 1087 1087
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT NP_BIND 1234 1241
FT SITE 1320 1323
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 299 299
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT CARBOHYD 1091 1091
FT CARBOHYD 2038 2038
FT CARBOHYD 2359 2359
FT CARBOHYD 2811 2811
SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FDA1 CRC64;

Query Match 45.0%; Score 9; DB 1; Length 3033;
Best Local Similarity 100.0%; Pred. No. 0.092;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYVGGVEHRL 9
Db 635 MYVGGVEHRL 643

FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).

FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).

FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).

FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).

FT CHAIN 730 1006 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).

FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).

FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).

FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).

FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).

FT TRANSMEM 347 369 POTENTIAL.

FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT NP_BIND 1230 1237 ATP (POTENTIAL).

FT SITE 1316 1319 DECH BOX.

FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).

SEQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64; .

Query Match 50.0%; Score 10; DB 1; Length 3010;

Best Local Similarity 100.0%; Pred. No. 0.0078;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYVGVVHRL 10

Db 631 MYVGVVHRL 640

RESULT 8

ID POLG_HCVTW STANDARD; PRT; 3010 AA.

AC P29846;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DE 15-SEP-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)).

OS Hepatitis C virus (isolate Taiwan) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=31645;

RP [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92230206; PubMed=1314449;

RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;

RT "The Taiwanese hepatitis C virus genome: sequence determination and mapping the 5' termini of viral genomic and antigenomic RNA.";

RL Virology 188:102-113(1992).

CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.

CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

-!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.

-!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).

-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M84754; -; NOT_ANNOTATED_CDS.

PIR; A40244; GNMVTM.

PDB; 1N64; 25-FEB-03.

PDB; 1NS3; 08-APR-98.

MEROPS; S29.001; -.

MEROPS; U39.001; -.

InterPro; IPR001410; DEAD.

InterPro; IPR002522; HCV_capsid.

InterPro; IPR002521; HCV_core.

InterPro; IPR002519; HCV_env.

InterPro; IPR002531; HCV_NS1.

InterPro; IPR002538; HCV_NS2.

InterPro; IPR004109; HCV_NS3.

InterPro; IPR000745; HCV_NS4a.

InterPro; IPR001490; HCV_NS4b.

InterPro; IPR002868; HCV_NS5a.

InterPro; IPR002166; HCV_RdRp.

InterPro; IPR007095; RNA_pol_DS_PS.

InterPro; IPR007094; RNA_pol_PSVir.

Pfam; PF01543; HCV_capsid; 1.

Pfam; PF01542; HCV_core; 1.

Pfam; PF01539; HCV_env; 1.

Pfam; PF01560; HCV_NS1; 1.

Pfam; PF01538; HCV_NS2; 1.

Pfam; PF02907; HCV_NS3; 1.

Pfam; PF01006; HCV_NS4a; 1.

Pfam; PF01001; HCV_NS4b; 1.

Pfam; PF01506; HCV_NS5a; 1.

Pfam; PF00271; helicase_C; 1.

Pfam; PF00998; Viral_RdRp; 1.

ProDom; PD186062; HCV_NS1; 1.

SMART; SM00487; DEXDC; 1.

Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase; Core protein; Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane; Nonstructural protein; Hydrolase; Serine protease; 3D-structure. 1 1

INIT_NET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE.

CHAIN 1 115 CORE PROTEIN (POTENTIAL).

CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).

CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).

CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).

CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).

CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).

CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).

CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).

CHAIN 347 369 TRANSMEM.

ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).

ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).

ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).

NP_BIND 1230 1237 ATP (POTENTIAL).

SITE 1316 1319 DECH BOX.

DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein C BY THE
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;
 Query Match 50.0%; Score 10; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.0078;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MYVGVGVEHRL 10
 DB 631 MYVGVGVEHRL 640
 RESULT 7
 POLG_HCVJUT STANDARD; PRT; 3010 AA.
 AC Q00269;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

OS Hepatitis C virus (isolate HC-JT) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92295714; PubMed=1318627;
 RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 RT "Molecular cloning of hepatitis C virus genome from a single Japanese
 RT carrier: sequence variation within the same individual and among
 RT infected individuals.";
 RL Virus Res. 23:39-53(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA) (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; D11168; BAA01943.1; -.
 DR PIR; A45573; A45573.
 DR PDB; 1A1Q; 25-MAR-98.
 DR PDB; 1JXP; 14-JAN-98.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR007095; RNA pol_PS ps.
 DR InterPro; IPR007094; RNA pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein C BY THE
 KW 3D-structure.
 FT INIT_MET 1 1
 FT REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.

FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2529 2529
 FT CARBOHYD 2788 2788
 FT STRAND 1031 1035
 FT HELIX 1039 1047
 FT STRAND 1050 1050
 FT STRAND 1059 1063
 FT STRAND 1068 1074
 FT TURN 1075 1076
 FT STRAND 1077 1081
 FT HELIX 1082 1085
 FT TURN 1086 1087
 FT STRAND 1090 1092
 FT TURN 1093 1094
 FT STRAND 1095 1097
 FT STRAND 1101 1103
 FT TURN 1104 1107
 FT STRAND 1108 1112
 FT STRAND 1122 1122
 FT STRAND 1129 1133
 FT TURN 1135 1136
 FT STRAND 1139 1144
 FT STRAND 1149 1157
 FT HELIX 1158 1161
 FT TURN 1162 1163
 FT TURN 1165 1166
 FT STRAND 1168 1171
 FT TURN 1172 1174
 FT STRAND 1175 1186
 FT TURN 1187 1188
 FT STRAND 1189 1197
 FT HELIX 1198 1202
 FT TURN 1203 1204
 FT STRAND 1680 1688
 SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCDFD9C CRC64;

Query Match 50.0%; Score 10; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.0078;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRL 10
 |||||
 Db 631 MYVGGVEHRL 640

RESULT 6
 POLG_HCVJA
 ID_POLG_HCVJA STANDARD; PRT; 3010 AA.
 AC P26662;

DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Japanese) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NCBI_TaxID=11116;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=91088550; PubMed=2175903;
 RX Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 RA Sugimura T., Shimotohno K.;
 RA "Molecular cloning of the human hepatitis C virus genome from
 RT Japanese patients with non-A, non-B hepatitis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RN [2]
 DISCUSSION OF SEQUENCE.
 MEDLINE=91192160; PubMed=1849488;
 RX Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K.,
 RA Ohkoshi S., Shimotohno K.;
 RA "Molecular structure of the Japanese hepatitis C viral genome.";
 RL FEBS Lett. 280:325-328(1991).
 CC -I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA}(N).
 CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: OF
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC
 CC -----
 CC EMBL; D90208; BAA14233.1; -.
 CC PIR; A39253; GNMVCJ.
 CC HSPP; P26663; IJXP.
 CC MEROPS; S29.001; -.
 CC MEROPS; U39.001; -.
 CC InterPro; IPR001410; DEAD.
 CC InterPro; IPR002522; HCV_capsid.
 CC InterPro; IPR002521; HCV_core.
 CC InterPro; IPR002519; HCV_env.
 CC InterPro; IPR002531; HCV_NS1.
 CC InterPro; IPR002518; HCV_NS2.
 CC InterPro; IPR004109; HCV_NS3.
 CC InterPro; IPR000745; HCV_NS4a.
 CC InterPro; IPR001490; HCV_NS4b.
 CC InterPro; IPR002868; HCV_NS5a.
 CC InterPro; IPR002166; HCV_RdRP.
 CC InterPro; IPR001650; Helicase_C.
 CC InterPro; IPR007095; RNA_pol_DS_PS.
 CC InterPro; IPR007094; RNA_pol_PSVir.
 CC Pfam; PF01543; HCV_capsid; 1.
 CC Pfam; PF01542; HCV_core; 1.
 CC Pfam; PF01539; HCV_env; 1.

FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 737 737
 SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122F2 CRC64;
 Query Match 50.0%; Score 10; DB 1; Length 737;
 Best Local Similarity 100.0%; Pred. No. 0.0023;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MYGVGVHRL 10
 DB 635 MYGVGVHRL 644
 RESULT 5
 POLG_HCVBK STANDARD; PRT; 3010 AA.
 AC P26663;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein (Contains: Capsid protein c (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate BK) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11105;
 RN [1]_TaxID=11105;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91140698; PubMed=1847440;
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 RA Onishi E., Andoh T., Yoshida I., Okayama H.;
 RT "Structure and organization of the hepatitis C virus genome isolated
 RT from human carriers.";
 RL J. Virol. 65:1105-1113 (1991).
 RN [2]
 RP SEQUENCE OF 1487-1500.
 RX MEDLINE=96235224; PubMed=8647104;
 RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetky L.;
 RT "Non-structural protein 3 of hepatitis C virus inhibits
 RT phosphorylation mediated by cAMP-dependent protein kinase.";
 RL Eur. J. Biochem. 237:611-618 (1996).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RX MEDLINE=97015088; PubMed=8861916;
 RA Love R.A., Farge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
 RA Moonaw E.W., Adachi T., Hostomsky Z.;
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 RT trypsin-like fold and a structural zinc binding site.";
 RL Cell 87:331-342 (1996).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RX MEDLINE=98227846; PubMed=9568891;
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
 RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
 RL Protein Sci. 7:837-847 (1998).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.

CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: Nucleoside triphosphate = N diphosphate +
 CC [RNA] (N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M58335; AAA72945.1; -;
 DR PIR; A38485; GNVWTC.
 DR PDB; 1AIQ; 25-MAR-98.
 DR PDB; 1JXP; 14-JAN-98.
 DR PDB; 1NS3; 08-APR-98.
 DR PDB; 1C2P; 15-NOV-00.
 DR PDB; 1CSJ; 08-NOV-99.
 DR PDB; 1GX5; 09-APR-02.
 DR PDB; 1GX6; 10-APR-02.
 DR PDB; 1QVU; 26-JUN-00.
 DR PDB; 8OHM; 20-APR-99.
 DR MEROPS; S29.001; -;
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRP.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXdc; 1.
 DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT CHAIN

DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
 DE protein (Envelope protein M); Major envelope protein E; Nonstructural
 DE proteins NS1 and NS2] (Fragment).
 OS Hepatitis C virus (isolate HC-J5) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NCBI_TaxID=11112;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL: D10075; BAA00969.1; -;
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR Pfam: PF01542; HCV capsid; 1.
 DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 DR Transmembrane; Nonstructural protein.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT NON TER 737 737
 SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CRC64;

Query Match

50.0%; Score 10; DB 1; Length 737;

Best Local Similarity 100.0%; Pred. No. 0.0023;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MYVGGVEHRL 10
 DB 635 MYVGGVEHRL 644
 RESULT 4
 POLG_HCVJ7
 ID POLG_HCVJ7 STANDARD; PRT; 737 AA.
 AC P27961;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
 DE protein (Envelope protein M); Major envelope protein E; Nonstructural
 DE proteins NS1 and NS2] (Fragment).
 OS Hepatitis C virus (isolate HC-J7) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NCBI_TaxID=11114;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL: D10077; BAA00971.1; -;
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR Pfam: PF01543; HCV capsid; 1.
 DR Pfam: PF01542; HCV capsid; 1.
 DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 DR Transmembrane; Nonstructural protein.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417

CC EMBL; M67463; AAA5534.1; --
DR FIR; A36814; GNVVCH.
DR PDB; 1HEI; 25-NOV-98.
DR PDB; 1AIV; 16-FEB-99.
DR PDB; 1AJR; 17-JUN-98.
DR MEROPS; S29.001; --
DR MEROPS; U39.001; --
DR TRANSFAC; T04155; --
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA pol Ds ps.
DR InterPro; IPR007094; RNA pol_psvir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; helicase C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
FT CHAIN 192 383 CAPSID PROTEIN C.
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
FT CHAIN 810 1026 PROTEIN P7.
FT CHAIN 1027 1637 NONSTRUCTURAL PROTEIN NS2.
FT CHAIN 1638 1711 PROTEASE/HELICASE NS3.
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4.
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5A.
FT CHAIN 347 369 NONSTRUCTURAL PROTEIN NS5B.
FT TRANSMEM 369 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).

FT STRAND 1224 1226
FT TURN 1232 1233
FT TURN 1236 1238
FT HELIX 1239 1246
FT TURN 1247 1248
FT STRAND 1251 1255
FT HELIX 1258 1271
FT TURN 1272 1272
FT STRAND 1277 1280
FT TURN 1281 1282
FT STRAND 1283 1285
FT STRAND 1291 1295
FT HELIX 1296 1301
FT TURN 1302 1303
FT STRAND 1312 1316
FT TURN 1317 1319
FT HELIX 1323 1335
FT TURN 1336 1340
FT STRAND 1343 1347
FT TURN 1352 1353
FT TURN 1360 1361
FT STRAND 1362 1366
FT STRAND 1368 1368
FT STRAND 1373 1375
FT TURN 1376 1377
FT STRAND 1378 1380
FT HELIX 1382 1385
FT STRAND 1389 1393
FT HELIX 1397 1409
FT TURN 1410 1411
FT STRAND 1414 1417
FT TURN 1419 1420
FT STRAND 1432 1436
FT TURN 1438 1439
FT STRAND 1450 1453
FT STRAND 1456 1463
FT STRAND 1471 1478
FT STRAND 1480 1480
FT HELIX 1481 1488
FT TURN 1489 1490
FT STRAND 1497 1501
FT STRAND 1507 1507
FT STRAND 1511 1511
FT HELIX 1514 1527
FT HELIX 1532 1544
FT STRAND 1550 1550
FT HELIX 1555 1564
FT HELIX 1570 1578
FT TURN 1579 1580
FT HELIX 1584 1597
FT TURN 1598 1598
FT HELIX 1606 1611
FT TURN 1614 1618
FT STRAND 1622 1623
FT STRAND 1627 1627
FT STRAND 1635 1636
FT HELIX 1640 1652
SQ SEQUENCE 3011 AA; 772CBB29CCD94753 CRC64;

Query Match 85.0%; Score 17; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.7e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGVEHRLAECNWT 17
Db 631 MYVGVEHRLAECNWT 647

RESULT 3
POLG_HCVJ5
ID POLG_HCVJ5 STANDARD; PRT; 737 AA.
AC P27960;
DT 01-AUG-1992 (Rel. 23, Created)

DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_NS5b.
 DR InterPro: IPR001650; Helicase C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase C; 1.
 DR Pfam: PF00998; Viral RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDc; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1616
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3011
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 476 476
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2364 2364
 FT CARBOHYD 2789 2789
 SEQUENCE 3011 AA; 327197 MW; 65F8C9447CE5AF9 CRC64;

Query Match 85.0%; Score 17; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 2.7e-10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYGVGVEHLEAACNWT 17

Db 631 MYGVGVEHLEAACNWT 647
 |||||
 POLG_HCVH STANDARD; PRT; 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 CC Hepacivirus.
 CC NCBI_TaxID=11108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=92052256; PubMed=1658800;
 RX Inchauspe G., Zebdee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154331; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 RL Structure 6:89-100(1998).
 CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1',
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC [RNA] (N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:52 ; Search time 4.85 Seconds
(without alignments)
193.925 Million cell updates/sec

Title: US-09-973-025-90
Perfect score: 20
Sequence: 1 MYVGVHRLEAACNWTGPE 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	85.0	3011	1 POLG_HCV1	P26864 h genome po
2	17	85.0	3011	1 POLG_HCVH	P27958 h genome po
3	10	50.0	737	1 POLG_HCVJ5	P27960 hepatitis c
4	10	50.0	737	1 POLG_HCVJ7	P27961 hepatitis c
5	10	50.0	3010	1 POLG_HCVBK	P26663 h genome po
6	10	50.0	3010	1 POLG_HCVJA	P26662 h genome po
7	10	50.0	3010	1 POLG_HCVJT	P26661 h genome po
8	10	50.0	3010	1 POLG_HCVJW	P26660 h genome po
9	10	50.0	3033	1 POLG_HCVJ6	P26659 h genome po
10	9	45.0	3033	1 POLG_HCVJ8	P26658 h genome po
11	7	35.0	518	1 T408_BURCE	Q51649 burkholderi
12	7	35.0	952	1 SYL_CORGL	Q8nlc4 corynebacte
13	7	35.0	957	1 SYL_COREF	Q8flm0 corynebacte
14	6	30.0	107	1 TBCA_RABIT	P80584 oryctolagus
15	6	30.0	168	1 DH1_MAIZE	P12950 zea mays [m
16	6	30.0	195	1 TSA1_YEAST	P34760 saccharomyc
17	6	30.0	195	1 TSA2_YEAST	Q04120 saccharomyc
18	6	30.0	265	1 DAPB_AQUAE	O67061 aquifex aeo
19	6	30.0	387	1 THIK_ECO57	Q8xj44 escherichia
20	6	30.0	387	1 THIK_ECOLI	P21151 escherichia
21	6	30.0	387	1 THIK_SALTY	Q91616 salmonella
22	6	30.0	390	1 THIK_PSEFR	P28790 pseudomonas
23	6	30.0	391	1 POLG_CVWV	P20177 clover yell
24	6	30.0	401	1 CSD_PSEPK	Q92408 pseudomonas
25	6	30.0	410	1 Y2B_MICTU	Q50700 mycobacteri
26	6	30.0	450	1 MURD_BACHD	Q9K988 bacillus ha
27	6	30.0	465	1 MURD_CHLTE	Q8kgd2 chlorobium
28	6	30.0	465	1 Y065_MYCPN	P75612 mycoplasma
29	6	30.0	516	1 Y4BL_RHISN	P55379 rhizobium s
30	6	30.0	535	1 YFHE_YEAST	P43390 saccharomyc
31	6	30.0	571	1 ILVI_BUCAP	O85293 buchnera ap
32	6	30.0	675	1 POLG_PRSVP	P19723 papaya ring
33	6	30.0	675	1 POLG_PRSVW	P19724 papaya ring

RESULT 1

ID	POLG_HCV1	STANDARD;	PRT;	3011 AA.
AC	P26664;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Genome polyprotein [Contains: Capsid protein C (core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].			
OS	Hepatitis C virus (isolate 1) (HCV).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.			
OX	NCBI_TaxID=11104;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91172826; PubMed=1848704;			
RA	Choo O.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.; "Genetic organization and diversity of the hepatitis C virus."; Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).			
RL	FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.			
CC	NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION. precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	-!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA}(N).			
CC	-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.			
CC	-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; M62321; AAA45676.1; -			
CC	PIR; A39166; GNMVC3.			
CC	PDB; 1A1V; 16-FEB-99.			
CC	PDB; 1HEI; 25-NOV-98.			
CC	MEROPS; S29.001; -			
CC	MEROPS; U39.001; -			
CC	InterPro; IPR001410; DEAD.			
CC	InterPro; IPR002522; HCV_capsid.			

34	6	30.0	870	1 YCS0_YEAST	P25623 saccharomyc
35	6	30.0	1485	1 CFTR_XENLA	P26363 xenopus lae
36	6	30.0	3061	1 POLG_PVYHU	Q02963 p genome po
37	6	30.0	3063	1 POLG_PVYN	P18247 p genome po
38	6	30.0	3344	1 POLG_PSVH	Q01801 p genome po
39	5	25.0	17	1 SP51_BACLI	P27642 bacillus li
40	5	25.0	31	1 CTRP_PENMO	P35002 penaeus mon
41	5	25.0	61	1 VGS2_BPMO2	O64242 mycobacteri
42	5	25.0	83	1 CP69_BP186	P08685 bacterioph
43	5	25.0	85	1 RUXF_CABEL	P34659 caenorhabdi
44	5	25.0	100	1 SECG_AQUAE	O66505 aquifex aeo
45	5	25.0	107	1 TBCA_BOVIN	P48427 bos taurus

ALIGNMENTS

THIS PAGE BLANK (USPTO)

RESULT 13

GNWVTW

genome polyprotein - hepatitis C virus (strain Taiwan)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain Taiwan)
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 A:Note: host Homo sapiens (man)
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C:Accession: A40244

R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
 Virology 188, 102-113, 1992
 A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
 A:Reference number: A40244; MUID:92230206; PMID:1314449

A:Accession: A40244
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <CHE>

A:Cross-references: GB:M84754
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEP>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome polyprotein NS5 #status predicted <NS5>
 F:1230-1317/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: DEXH motif
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,233,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 50.0%; Score 10; DB 1; Length 3010;

Best Local Similarity 100.0%; Pred. No. 0.018;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRL 10

Db 631 MYVGGVEHRL 640

RESULT 14

S40770

genome polyprotein - hepatitis C virus
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain Taiwan)
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus

C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001

C:Accession: S40770; PC1285

R:Okamoto, H.
 submitted to the EMBL Data Library, March 1992

A:Reference number: S40770

A:Accession: S40770

A:Molecule type: genomic RNA

A:Residues: 1-3011 <OKA>

A:Cross-references: EMBL:D10749; NID:G221586; PIDN:BAA01582.1; PID:G221587

Jpn. J. Exp. Med. 60, 167-177, 1990
 A:Title: The 5'-terminal sequence of the hepatitis C virus genome.

A:Reference number: PC1284; MUID:91013116; PMID:2170712

A:Accession: PC1285

A:Molecule type: genomic RNA

A:Residues: 1-513 <OK2>

A:Cross-references: GB:D00831; NID:G221511; PIDN:BAA00705.1; PID:G221512

A:Experimental source: isolate HC-J1
 C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin

F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: major envelope protein E #status predicted <MEP>

F:192-389/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome polyprotein NS5 #status predicted <NS5>
 F:1230-1317/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: DEXH motif
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,233,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome polyprotein NS5 #status predicted <NS5>
 F:1230-1317/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: DEXH motif
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 50.0%; Score 10; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 0.018;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRL 10

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Db 631 MYVGGVEHRL 640

Search completed: November 21, 2003, 21:11:39
 Job time : 10.25 secs

RESULT 10

GNVVCJ

genome polyprotein - hepatitis C virus (strain J)
 N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
 C:Accession: A39253; PS0086
 R:Kato, N.; Hijikata, M.; Ootsubawa, Y.; Nakagawa, M.; Sugimura, T.; Shimotohno, R.; Kato, N.; Hijikata, M.; U.S.A. 87, 9524-9528, 1990
 Proc. Natl. Acad. Sci. U.S.A.
 A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients
 A:Reference number: A39253; MUID:91088550; PMID:2175903
 A:Accession: A39253

A:Molecule type: Genomic RNA

A:Residues: 1-3010 <KAT>

A:Cross-references: GB:D90208; NID:G221610; PIDN:BAA14233.1; PID:G221611

R:Kato, N.; Ohkoshi, S.; Shimotohno, K.

Proc. Jpn. Acad. 55B, 219-223, 1989

A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence variation

A:Reference number: PS0085

A:Accession: PS0086

A:Molecule type: Genomic RNA

A:Residues: 2650-2707 <KA2>

A:Experimental source: Japanese isolate

C:Comment: The cleavage sites of this polyprotein have not been determined.

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine

F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitis C virus genome polyprotein

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:196,209,234,250,305,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2400

F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome polyprotein
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 50.0%; Score 10; DB 1; Length 3010;

Best Local Similarity 100.0%; Pred. No. 0.018;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRL 10

|||||

Db 631 MYVGGVEHRL 640

RESULT 12

S18030

genome polyprotein - hepatitis C virus (isolate JK1)

N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein NS5

C:Species: hepatitis C virus

A:Variety: isolate JK1

C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001

C:Accession: S18030; S33570; A48332; S18029

R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.

submitted to the EMBL Data Library, September 1991

A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient

A:Reference number: S18028

A:Accession: S18030

A:Molecule type: Genomic RNA

A:Residues: 1-3010 <HON>

A:Cross-references: EMBL:X61596; NID:G59478; PIDN:CAA43793.1; PID:G59479

A:Experimental source: isolate JK1 from an individual

R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.

Arch. Virol. 128, 163-169, 1993

A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolated

A:Reference number: A48332; MUID:93119270; PMID:8380322

A:Accession: S33570

A:Molecule type: Genomic RNA

A:Residues: 1-547, 'I', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HON>

A:Cross-references: EMBL:X61591

A>Note: this sequence is inconsistent with the nucleotide translation

A>Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320

as Trp, and TTC for residue 771 as Ser

A>Note: sequence extracted from NCBI backbone (NCBI:121747, NCBI:121748)

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine

F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitis C virus genome polyprotein

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

F:192-389/Product: envelope protein E1 #status predicted <E1>
 F:390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predicted <N1>
 F:196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn) (cova

Query Match 50.0%; Score 10; DB 2; Length 640;
 Best Local Similarity 100.0%; Pred. No. 0.0046;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRL 10
 |||||
 Db 631 MYVGGVEHRL 640

RESULT 6

JQ1366
 genome polyprotein - hepatitis C virus (French isolate) (fragments)
 C:Species: hepatitis C virus
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
 C:Accession: JQ1366
 R:Krensdorff, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.
 J. Gen. Virol. 72, 2557-2561, 1991
 A:Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication
 A:Reference number: JQ1366; MUID:92013977; PMID:1655961
 A:Accession: JQ1366
 A:Molecule type: genomic RNA
 A:Residues: 1-716 <KRE>
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: glycoprotein; polyprotein
 F:84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #stat

Query Match 50.0%; Score 10; DB 2; Length 716;
 Best Local Similarity 100.0%; Pred. No. 0.005;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRL 10
 |||||
 Db 298 MYVGGVEHRL 307

RESULT 7

S19876
 genome polyprotein - hepatitis C virus (isolate JK5) (fragment)
 N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
 C:Species: hepatitis C virus
 A:Variety: isolate JK5
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
 C:Accession: S19876
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso

A:Accession: S19876
 A:Molecule type: genomic RNA
 A:Residues: 1-782 <HON>

A:Cross-references: EMBL:X61595; NID:g59486; PIDN:CAA43792.1; PID:g59487
 A:Experimental source: isolate JK5

C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
 F:1-191/Product: core protein #status predicted <N1>
 F:192-383/Product: envelope protein 1 #status predicted <MAT2>
 F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
 F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 50.0%; Score 10; DB 2; Length 782;
 Best Local Similarity 100.0%; Pred. No. 0.0054;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRL 10
 |||||
 Db 631 MYVGGVEHRL 640

RESULT 8

S18031

genome polyprotein - hepatitis C virus (isolate JK2) (fragment)
 N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
 C:Species: hepatitis C virus
 A:Variety: isolate JK2
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
 C:Accession: S18031
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus is

A:Reference number: S18029
 A:Accession: S18031
 A:Molecule type: genomic RNA

A:Residues: 1-782 <HON>
 A:Cross-references: EMBL:X61593

A:Experimental source: isolate JK2
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural

F:1-191/Product: core protein #status predicted <MAT1>
 F:192-383/Product: envelope protein 1 #status predicted <MAT2>
 F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
 F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 50.0%; Score 10; DB 2; Length 782;
 Best Local Similarity 100.0%; Pred. No. 0.0054;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRL 10
 |||||
 Db 631 MYVGGVEHRL 640

RESULT 9

GNVTC

genome polyprotein - hepatitis C virus
 N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstruct
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
 C:Accession: A38465
 R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.;
 J. Virol. 65, 1105-1113, 1991
 A:Title: Structure and organization of the hepatitis C virus genome isolated from human

A:Reference number: A38465; MUID:91140698; PMID:1847440
 A:Accession: A38465
 A:Molecule type: genomic RNA

A:Residues: 1-3010 <TAK>
 A:Cross-references: EMBL:M58335; NID:g329770; PIDN:AAA72945.1; PID:g329771

C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructur

F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-723/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: hepacivirin #status predicted <NS2>
 F:1007-1615/Product: hepacivirin #status predicted <NS3>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,2;
 Query Match 50.0%; Score 10; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.018;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRL 10
 |||||
 Db 631 MYVGGVEHRL 640

A;Reference number: PQ0393; MUID:92268871; PMID:1316939

A;Accession: PQ0403

A;Molecule type: genomic RNA

A;Residues: 1577-1633 <CHA>

A;Cross-references: DDBJ:D10128

A;Experimental source: isolates E-b16

A;Accession: PQ0404

A;Status: preliminary

A;Molecule type: genomic RNA

A;Residues: 1577-1633 <CH2>

A;Experimental source: isolates E-b17

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural

F;1-115/Product: capsid protein C #status predicted <CPC>

F;116-191/Product: envelope protein M #status predicted <EPM>

F;192-389/Product: major envelope protein E #status predicted <MEE>

F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F;1007-1615/Product: hepatitis C virus genome polyprotein

F;1230-1237/Region: nucleotide-binding motif A (P-loop)

F;1312-1317/Region: nucleotide-binding motif B

F;1316-1319/Region: DEXH motif

F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 85.0%; Score 17; DB 1; Length 3011;

Best Local Similarity 100.0%; Pred. No. 6.2e-10;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGVGVEHRLAACNWT 17

|||||

631 MYVGVGVEHRLAACNWT 647

RESULT 3

GNMVCH

genome polyprotein - hepatitis C virus (strain H)

N;Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; seri

F;1-115/Product: capsid protein C #status predicted <CPC>

F;116-191/Product: envelope protein M #status predicted <EPM>

F;192-389/Product: major envelope protein E #status predicted <MEE>

F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F;1007-1615/Product: hepatitis C virus genome polyprotein

F;1230-1237/Region: nucleotide-binding motif A (P-loop)

F;1312-1317/Region: nucleotide-binding motif B

F;1316-1319/Region: DEXH motif

F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,2

Query Match 85.0%; Score 17; DB 1; Length 3011;

Best Local Similarity 100.0%; Pred. No. 6.2e-10;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGVGVEHRLAACNWT 17

|||||

631 MYVGVGVEHRLAACNWT 647

RESULT 4

JCS620

genome polyprotein - hepatitis C virus (isolate EUH1480)

N;Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; seri

F;1-115/Product: capsid protein C #status predicted <CPC>

F;116-191/Product: envelope protein M #status predicted <EPM>

F;192-389/Product: major envelope protein E #status predicted <MEE>

F;384-408/Region: hypervariable #status predicted

F;390-730/Product: nonstructural protein NS1 #status predicted <NS1>

F;731-1007/Product: nonstructural protein NS2 #status predicted <NS2>

F;1008-1616/Product: hepatitis C virus genome polyprotein

F;1231-1238/Region: nucleotide-binding motif A (P-loop)

F;1313-1318/Region: nucleotide-binding motif B

F;1317-1320/Region: DEXH motif

F;1617-1863/Product: nonstructural protein NS4a #status predicted <N4A>

F;1864-2014/Product: nonstructural protein NS4b #status predicted <N4B>

F;2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>

F;2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 55.0%; Score 11; DB 1; Length 3014;

Best Local Similarity 100.0%; Pred. No. 0.0015;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 EHRLEAACNWT 17

|||||

638 EHRLEAACNWT 648

RESULT 5

JQ1584

genome polyprotein - hepatitis C virus (strain U.K.) (fragment)

N;Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural pro

C;Species: hepatitis C virus

C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000

C;Accession: JQ1584

R;Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.

J. Gen. Virol. 73, 1521-1525, 1992

A;Title: Cloning and sequencing of the structural region and expression of putative cor

A;Reference number: JQ1584; MUID:92300349; PMID:1318944

A;Accession: JQ1584

A;Molecule type: genomic RNA

A;Residues: 1-640 <KUM>

A;Cross-references: GB:X84079; NID:G643119; PIDN:CAAS8888.1; PID:G643120

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; polypr

F;1-191/Product: core protein C #status predicted <CPC>

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:46:27 ; Search time 9.25 Seconds
(without alignments)
207.932 Million cell updates/sec

Title: US-09-973-025-90

Perfect score: 20

Sequence: 1 MYVGVGVEHRLAACNWTGCE 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR 76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	85.0	782	2	S18032 genome polyprotein
2	17	85.0	3011	1	GNMVC3 genome polyprotein
3	17	85.0	3011	1	GNMVCH genome polyprotein
4	11	55.0	3014	1	JC3620 genome polyprotein
5	10	50.0	640	2	JQ1584 genome polyprotein
6	10	50.0	716	2	JQ1366 polyprotein - hepa
7	10	50.0	782	2	S19876 genome polyprotein
8	10	50.0	782	2	S18031 genome polyprotein
9	10	50.0	3010	1	GNMVTC genome polyprotein
10	10	50.0	3010	1	GNMVCH genome polyprotein
11	10	50.0	3010	1	A45573 genome polyprotein
12	10	50.0	3010	1	S18030 genome polyprotein
13	10	50.0	3010	1	GNMVTV genome polyprotein
14	10	50.0	3011	1	S40770 genome polyprotein
15	10	50.0	3033	1	JQ1303 genome polyprotein
16	9	45.0	782	2	S19875 genome polyprotein
17	9	45.0	3033	1	GNMVJ8 genome polyprotein
18	7	35.0	192	2	C87670 hypothetical prote
19	7	35.0	206	2	AF3433 ribosomal-protein-
20	7	35.0	350	2	S35631 genome polyprotein
21	7	35.0	586	2	PN0677 phosphoprotein pho
22	7	35.0	787	2	FN0677 hypothetical prote
23	7	35.0	1367	2	T21913 hypothetical prote
24	6	30.0	167	2	S05545 dehydrin 3 - maize
25	6	30.0	168	2	A93116 RAB-17 protein - m
26	6	30.0	168	2	S08633 RAB-17 protein - m
27	6	30.0	174	2	H83481 hypothetical prote
28	6	30.0	196	2	A47362 thiol-specific ant
29	6	30.0	196	2	S69732 hypothetical prote

30 6 30.0 214 2 T51061 hypothetical prote
31 6 30.0 233 2 AF3522 transcription regu
32 6 30.0 265 2 B70379 dihydrodipicolinat
33 6 30.0 266 2 A70345 hypothetical prote
34 6 30.0 268 2 S61209 aminoglycoside pho
35 6 30.0 268 2 D83131 aminoglycoside 3'-
36 6 30.0 333 2 F83331 ABC transporter AT
37 6 30.0 348 2 T35450 genome polyprotein
38 6 30.0 379 2 S13556 acetyl-CoA C-acylt
39 6 30.0 387 1 XUEC fatty oxidation co
40 6 30.0 387 2 A82036 acetyl-CoA transfe
41 6 30.0 387 2 E91225 acetyl-CoA transfe
42 6 30.0 387 2 C86072 small (beta) chain
43 6 30.0 387 2 A10914 fatty-acid beta-ox
44 6 30.0 391 1 JS0624 genome polyprotein
45 6 30.0 393 2 JQ0461

ALIGNMENTS

RESULT 1

S18032 genome polyprotein - hepatitis C virus (isolate JK4) (fragment)
N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C:Species: hepatitis C virus
A:Variety: isolate JK4
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C:Accession: S18032
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
Submitted to the EMBL Data Library, September 1991
A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus isa
A:Reference number: S18029
A:Accession: S18032
A:Molecule type: Genomic RNA
A:Residues: 1-782 <HON>
A:Cross-references: EMBL:X61594
A:Experimental source: isolate JK4
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F:1-191/Product: core protein #status predicted <MAT1>
F:192-383/Product: envelope protein 1 #status predicted <MAT2>
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match

Best Local Similarity 85.0%; Score 17; DB 2; Length 782;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MYVGVGVEHRLAACNWT 17

Db 631 MYVGVGVEHRLAACNWT 647

RESULT 2

GNMVC3

genome polyprotein - hepatitis C virus (strain HCV-1)

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructu

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001

C:Accession: A39166; PQ0403; PQ0404

R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Co

Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991

A:Title: Genetic organization and diversity of the hepatitis C virus.

A:Reference number: A39166; MUID:91172826; PMID:1848704

A:Accession: A39166

A:Molecule type: mRNA

A:Residues: 1-3011 <CHO>

A:Cross-references: GB:M62321; NID:G329873; PIDN:AAA45676.1; PID:G329874

R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.F

J. Gen. Virol. 73, 1131-1141, 1992

A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to

QY 1 MYVGGVEHRLLEAA 13
Db 18 MYVGGVEHRLLEAA 30

RESULT 12
US-10-296-734-814
; Sequence 814, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; PRIOR FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU P07761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 814
; LENGTH: 2010
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC cassette B
US-10-296-734-814

Query Match 65.0%; Score 13; DB 6; Length 2010;
Best Local Similarity 100.0%; Pred. No. 5.9e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLLEAA 13
Db 78 MYVGGVEHRLLEAA 90

RESULT 13
US-09-350-841A-1217
; Sequence 1217, Application US/09350841A
; GENERAL INFORMATION:
; APPLICANT: Jeffs, Peter;
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION OF MEMBRANE
; FILE REFERENCE: 7872-066-999
; CURRENT APPLICATION NUMBER: US/09/350,841A
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 1946
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1217
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-09-350-841A-1217

Query Match 60.0%; Score 12; DB 5; Length 32;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YVGGVEHRLLEAA 13
Db 1 YVGGVEHRLLEAA 12

RESULT 14
US-10-351-641-1217
; Sequence 1217, Application US/10351641
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.

; APPLICANT: Lambert, D.
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; FILE REFERENCE: 7872-100
; CURRENT APPLICATION NUMBER: US/10/351,641
; PRIOR FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: 09/350,641
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/315,304
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1757
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1217
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-10-351-641-1217

Query Match 60.0%; Score 12; DB 6; Length 32;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YVGGVEHRLLEAA 13
Db 1 YVGGVEHRLLEAA 12

RESULT 15
US-09-350-841A-1904
; Sequence 1904, Application US/09350841A
; GENERAL INFORMATION:
; APPLICANT: Jeffs, Peter;
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION OF MEMBRANE
; FILE REFERENCE: 7872-066-999
; CURRENT APPLICATION NUMBER: US/09/350,841A
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 1946
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1904
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
; NAME/KEY: SITE
; LOCATION: 13, 21
; OTHER INFORMATION: Xaa = U (Aminobutyric Acid)
US-09-350-841A-1904

Query Match 60.0%; Score 12; DB 5; Length 34;
Best Local Similarity 100.0%; Pred. No. 1.7e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YVGGVEHRLLEAA 13
Db 1 YVGGVEHRLLEAA 12

Search completed: November 21, 2003, 22:12:59
Job time : 8.55 secs

```
; Sequence 406, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 406
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a consensus polyprotein
US-10-296-734-406

Query Match      85.0%; Score 17; DB 6; Length 3011;
Best Local Similarity 100.0%; Pred. No. 4.6e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MYVGGVEHRLAACNWT 17
DB      631 MYVGGVEHRLAACNWT 647
|||||
RESULT 8
US-10-296-734-810
; Sequence 810, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 810
; LENGTH: 5985
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a savine
US-10-296-734-810

Query Match      85.0%; Score 17; DB 6; Length 5985;
Best Local Similarity 100.0%; Pred. No. 8.6e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MYVGGVEHRLAACNWT 17
DB      1743 MYVGGVEHRLAACNWT 1759
|||||
RESULT 9
US-09-350-369D-1704
; Sequence 1704, Application US/09350369D
; GENERAL INFORMATION:
; APPLICANT: Samuel E. Hopkins
; APPLICANT: Elizabeth M. Dimassimo
; APPLICANT: Pamela P. RUSHAK
; APPLICANT: Thomas Michael Venetta
; TITLE OF INVENTION: Methods and Compositions for Detection of HIV GP41 Peptides and
; FILE REFERENCE: 7872-068-999
; CURRENT APPLICATION NUMBER: US/09/350,369D
; CURRENT FILING DATE: 1999-07-09
```

```
; NUMBER OF SEQ ID NOS: 2088
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1704
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus-1 (HIV-1)
US-09-350-369D-1704

Query Match      65.0%; Score 13; DB 5; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 GVEHRLAACNWT 17
DB      1 GVEHRLAACNWT 13
|||||
RESULT 10
US-09-350-369E-1704
; Sequence 1704, Application US/09350369E
; GENERAL INFORMATION:
; APPLICANT: Samuel E. Hopkins
; APPLICANT: Elizabeth M. Dimassimo
; APPLICANT: Pamela P. RUSHAK
; APPLICANT: Thomas Michael Venetta
; TITLE OF INVENTION: Methods and Compositions for Detection of HIV GP41 Peptides
; FILE REFERENCE: 7872-068-999
; CURRENT APPLICATION NUMBER: US/09/350,369E
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 2088
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1704
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus-1 (HIV-1)
US-09-350-369E-1704

Query Match      65.0%; Score 13; DB 5; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 GVEHRLAACNWT 17
DB      1 GVEHRLAACNWT 13
|||||
RESULT 11
US-10-296-734-490
; Sequence 490, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 490
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 42
US-10-296-734-490

Query Match      65.0%; Score 13; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.55 Seconds
(without alignments)
117.536 Million cell updates/sec

Title: US-09-973-025-90

Perfect score: 20
Sequence: 1 MYVGVHRLEAACNWTPE 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA_New.*

- 1: /cgn2_6/prodata/1/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/prodata/1/paa/US05_NEW_COMB.pep.*
- 3: /cgn2_6/prodata/1/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/prodata/1/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/prodata/1/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/prodata/1/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/prodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	85.0	30	6	US-10-296-734-492
2	17	85.0	333	1	PCT-US03-19834-3
3	17	85.0	333	6	US-10-655-562-4
4	17	85.0	637	1	PCT-US03-33610-4
5	17	85.0	2011	6	US-10-296-734-812
6	17	85.0	3011	1	PCT-US03-19834-2
7	17	85.0	3011	6	US-10-296-734-406
8	17	85.0	5985	6	US-10-296-734-810
9	13	65.0	30	5	US-09-350-369D-1704
10	13	65.0	30	5	US-09-350-369E-1704
11	13	65.0	30	6	US-10-296-734-490
12	13	65.0	2010	6	US-10-296-734-814
13	12	60.0	32	5	US-09-350-841A-1217
14	12	60.0	32	6	US-10-351-641-1217
15	12	60.0	34	5	US-09-350-841A-1904
16	12	60.0	34	5	US-09-350-369D-1828
17	12	60.0	34	6	US-09-350-369E-1828
18	12	60.0	34	6	US-10-351-641-1715
19	11	55.0	37	5	US-09-350-841A-1216
20	11	55.0	37	6	US-10-351-641-1216
21	11	55.0	39	5	US-09-350-841A-1903
22	11	55.0	39	5	US-09-350-369D-1827
23	11	55.0	39	5	US-09-350-369E-1827
24	11	55.0	39	6	US-10-351-641-1714
25	10	50.0	347	6	US-10-664-391-9
26	10	50.0	539	6	US-10-664-391-11

27	10	50.0	1026	1	PCT-US03-20409-3	Sequence 3, Appli
28	10	50.0	2280	1	PCT-US03-20322-211	Sequence 211, App
29	10	50.0	3033	6	US-10-009-002-5	Sequence 5, Appli
30	9	45.0	10	1	PCT-US03-31303-367	Sequence 367, App
31	9	45.0	10	6	US-10-677-754-367	Sequence 367, App
32	9	45.0	28	6	US-10-685-435-30	Sequence 30, Appl
33	9	45.0	107	1	PCT-US03-31303-247	Sequence 247, App
34	9	45.0	107	6	US-10-677-754-247	Sequence 247, App
35	9	45.0	130	1	PCT-US03-31303-243	Sequence 243, App
36	9	45.0	130	6	US-10-677-754-243	Sequence 243, App
37	9	45.0	255	1	PCT-US03-31303-233	Sequence 233, App
38	9	45.0	255	6	US-10-677-754-233	Sequence 233, App
39	9	45.0	281	1	PCT-US03-31303-235	Sequence 235, App
40	9	45.0	281	6	US-10-677-754-235	Sequence 235, App
41	9	45.0	308	1	PCT-US03-31303-249	Sequence 249, App
42	9	45.0	308	1	PCT-US03-31303-251	Sequence 251, App
43	9	45.0	308	6	US-10-677-754-249	Sequence 249, App
44	9	45.0	308	6	US-10-677-754-251	Sequence 251, App
45	8	40.0	20	6	US-10-651-165-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1
US-10-296-734-492
; Sequence 492, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 492
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 43
US-10-296-734-492

Query Match 85.0%; Score 17; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 6.9e-12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGVHRLEAACNWT 17
|||||
Db 3 MYVGVHRLEAACNWT 19
|||||

RESULT 2
PCT-US03-19834-3
; Sequence 3, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; FILE REFERENCE: IOWA:045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3

US-08-444-112-56
; Sequence 56, Application US/08444112
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,112
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 141 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-444-112-56

Query Match 85.0%; Score 17; DB 8; Length 141;
Best Local Similarity 100.0%; Pred. No. 1.6e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYVGGVEHRLAEACNWT 17
Db 6 MYVGGVEHRLAEACNWT 22

RESULT 15
US-08-403-590B-77
; Sequence 77, Application US/08403590B
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590B
; FILING DATE: 14-MAR-1995

; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 179 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-403-590B-77

Query Match 85.0%; Score 17; DB 8; Length 179;
Best Local Similarity 100.0%; Pred. No. 2e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYVGGVEHRLAEACNWT 17
Db 121 MYVGGVEHRLAEACNWT 137

Search completed: November 21, 2003, 22:09:54
Job time : 163.1 secs

```

Query Match      85.0%;   Score 17;   DB 8;   Length 39;
Best Local Similarity 100.0%;   Pred. No. 4.9e-10;
Matches 17;   Conservative 0;   Mismatches 0;   Indels
QY      1  MYVGGVEHRLEAACNWT 17
          |||||
Db      7  MYVGGVEHRLEAACNWT 23

```

US-08-403-590B-56
; Sequence 56, Application US/08403590B
; GENERAL INFORMATION:

APPLICANT: Cullen, David L.
APPLICANT: Rutter, William J.
TITLE OF INVENTION: NANV Diagnostics and Vaccines
NUMBER OF SEQUENCES: 777
CORRESPONDENCE ADDRESS:
CORRESPONDENCE ADDRESS:

ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,590B
FILING DATE: 14-MAR-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa B.

```

NAME: HAROLD, ALISA A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0110.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508)359-3876
TELEFAX: (508)359-3885
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 141 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-403-590B-56

```

```

;
; Length: 141 amino acids
;
; TYPE: amino acid
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
US-08-403-590B-56

Query Match      85.0%; Score 17; DB 8; Length 141;
Best Local Similarity 100.0%; Pred. No. 1.6e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0

```

RESULT 14

;;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,482
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 17-June-1994
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 39 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
; US-08-475-482-47

Query Match 85.0%; Score 17; DB 8; Length 39;
Best Local Similarity 100.0%; Pred. No. 4.9e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYVGVGVEHRLAACNWT 17
Db 7 MYVGVGVEHRLAACNWT 23

RESULT 10
US-08-477-072-47
; Sequence 47, Application US/08477072
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,482
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 07-June-1995
; APPLICATION NUMBER: 07/719,819

;;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,072
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 17-June-1994
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 39 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
; US-08-477-072-47

Query Match 85.0%; Score 17; DB 8; Length 39;
Best Local Similarity 100.0%; Pred. No. 4.9e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYVGVGVEHRLAACNWT 17
Db 7 MYVGVGVEHRLAACNWT 23

RESULT 11
US-08-477-582-47
; Sequence 47, Application US/08477582
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,582
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/262,037
; FILING DATE: 07-June-1995
; APPLICATION NUMBER: 07/719,819

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136

```

APPLICATION NUMBER: US/09/899,303
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 90:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 90:
US-09-899-303-90

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-13; 0; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLAACNWTPE 20
|||||
DB 1 MYVGGVEHRLAACNWTPE 20

RESULT 2

US-09-899-303A-90
Sequence 90, Application US/09899303A
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303A
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 90:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 90:
US-09-899-303A-90

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-13; 0; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MYVGGVEHRLAACNWTPE 20
|||||
DB 1 MYVGGVEHRLAACNWTPE 20

RESULT 3

US-09-973-025-90
Sequence 90, Application US/09973025
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/973,025
FILING DATE: 10-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 90:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 90:
US-09-973-025-90

Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-13; 0; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLAACNWTPE 20
|||||
DB 1 MYVGGVEHRLAACNWTPE 20

RESULT 4

US-09-995-791-90

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:58:12 ; Search time 163.1 Seconds
(without alignments)
111.578 Million cell updates/sec

Title: US-09-973-025-90

Perfect score: 20

Sequence: 1 MYVGVHRLEAACNWTGGE 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 5728757 seqs, 909918778 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA_Main:*

- 1: /cgn2_6/ptodata/1/paa/PCTUS COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06 COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07 COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08 COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US081 COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US082 COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US083 COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US084 COMB.pep.*
- 9: /cgn2_6/ptodata/1/paa/US085 COMB.pep.*
- 10: /cgn2_6/ptodata/1/paa/US086 COMB.pep.*
- 11: /cgn2_6/ptodata/1/paa/US087 COMB.pep.*
- 12: /cgn2_6/ptodata/1/paa/US088 COMB.pep.*
- 13: /cgn2_6/ptodata/1/paa/US089 COMB.pep.*
- 14: /cgn2_6/ptodata/1/paa/US090 COMB.pep.*
- 15: /cgn2_6/ptodata/1/paa/US091 COMB.pep.*
- 16: /cgn2_6/ptodata/1/paa/US092 COMB.pep.*
- 17: /cgn2_6/ptodata/1/paa/US093 COMB.pep.*
- 18: /cgn2_6/ptodata/1/paa/US094 COMB.pep.*
- 19: /cgn2_6/ptodata/1/paa/US095 COMB.pep.*
- 20: /cgn2_6/ptodata/1/paa/US096 COMB.pep.*
- 21: /cgn2_6/ptodata/1/paa/US097A COMB.pep.*
- 22: /cgn2_6/ptodata/1/paa/US097B COMB.pep.*
- 23: /cgn2_6/ptodata/1/paa/US098 COMB.pep.*
- 24: /cgn2_6/ptodata/1/paa/US099A COMB.pep.*
- 25: /cgn2_6/ptodata/1/paa/US099B COMB.pep.*
- 26: /cgn2_6/ptodata/1/paa/US100 COMB.pep.*
- 27: /cgn2_6/ptodata/1/paa/US101 COMB.pep.*
- 28: /cgn2_6/ptodata/1/paa/US102 COMB.pep.*
- 29: /cgn2_6/ptodata/1/paa/US103 COMB.pep.*
- 30: /cgn2_6/ptodata/1/paa/US104 COMB.pep.*
- 31: /cgn2_6/ptodata/1/paa/US106 COMB.pep.*
- 32: /cgn2_6/ptodata/1/paa/US60 COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	20	100.0	20	23	US-09-899-303-90

2	20	100.0	20	23	US-09-899-303A-90	Sequence 90, Appl
3	20	100.0	20	25	US-09-973-025-90	Sequence 90, Appl
4	20	100.0	20	25	US-09-995-791-90	Sequence 90, Appl
5	20	100.0	20	25	US-09-995-808-90	Sequence 90, Appl
6	20	100.0	20	25	US-09-995-860-90	Sequence 90, Appl
7	20	100.0	20	26	US-10-020-510-90	Sequence 90, Appl
8	20	100.0	20	29	US-10-321-798-90	Sequence 90, Appl
9	17	85.0	39	8	US-08-475-482-47	Sequence 47, Appl
10	17	85.0	39	8	US-08-477-072-47	Sequence 47, Appl
11	17	85.0	39	8	US-08-477-582-47	Sequence 47, Appl
12	17	85.0	39	8	US-08-480-253-47	Sequence 47, Appl
13	17	85.0	141	8	US-08-403-590B-56	Sequence 56, Appl
14	17	85.0	141	8	US-08-444-112-56	Sequence 56, Appl
15	17	85.0	179	8	US-08-403-590B-77	Sequence 77, Appl
16	17	85.0	179	8	US-08-444-112-77	Sequence 77, Appl
17	17	85.0	250	1	PCT-US01-28767-8	Sequence 8, Appl
18	17	85.0	250	25	US-09-952-572-8	Sequence 8, Appl
19	17	85.0	254	1	PCT-US00-26395-3	Sequence 3, Appl
20	17	85.0	254	18	US-09-407-430-3	Sequence 3, Appl
21	17	85.0	278	8	US-08-471-498-15	Sequence 15, Appl
22	17	85.0	333	8	US-08-417-478-12	Sequence 12, Appl
23	17	85.0	333	8	US-08-453-613-12	Sequence 12, Appl
24	17	85.0	333	30	US-10-445-724-3	Sequence 3, Appl
25	17	85.0	333	32	US-60-409-909-4	Sequence 4, Appl
26	17	85.0	350	21	US-09-705-547-4	Sequence 4, Appl
27	17	85.0	350	24	US-09-929-955-4	Sequence 4, Appl
28	17	85.0	350	27	US-10-104-966-4	Sequence 4, Appl
29	17	85.0	350	32	US-60-229-175-4	Sequence 4, Appl
30	17	85.0	363	1	PCT-US00-26395-2	Sequence 2, Appl
31	17	85.0	363	18	US-09-407-430-2	Sequence 2, Appl
32	17	85.0	363	27	US-10-128-587A-97	Sequence 97, Appl
33	17	85.0	363	27	US-10-128-590-97	Sequence 97, Appl
34	17	85.0	367	8	US-08-417-478-6	Sequence 6, Appl
35	17	85.0	367	8	US-08-453-613-6	Sequence 6, Appl
36	17	85.0	399	8	US-08-417-478-10	Sequence 10, Appl
37	17	85.0	399	8	US-08-453-613-10	Sequence 10, Appl
38	17	85.0	409	8	US-08-471-498-21	Sequence 21, Appl
39	17	85.0	409	8	US-08-471-498-24	Sequence 24, Appl
40	17	85.0	414	3	US-07-748-561-2	Sequence 2, Appl
41	17	85.0	414	3	US-07-989-843-32	Sequence 32, Appl
42	17	85.0	414	8	US-08-463-849-32	Sequence 32, Appl
43	17	85.0	414	8	US-08-463-884-32	Sequence 32, Appl
44	17	85.0	414	20	US-09-690-359-32	Sequence 32, Appl
45	17	85.0	421	29	US-10-365-620-70	Sequence 70, Appl

ALIGNMENTS

RESULT 1

US-09-899-303-90
; Sequence 90, Application US/09899303
; GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

BOSMAN, FONS

DE MARTYNOFF, GUY

BUYSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHUYE P.C.

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

THIS PAGE BLANK (USPTO)

Db 459 MYVGVGVEHLEAACNWT 475
|||||

RESULT 15
US-09-941-611-23
; Sequence 23, Application US/09941611
; Patent No. US20020106640A1
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; POLLET, DIRK
; MAERTENS, GEERT
; VAN HEUVERSWUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; ANTIODIES TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/941,611
; APPLICATION NUMBER: US/09/941,611
; FILING DATE: 30-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/391,671
; FILING DATE: 1995-02-21
; APPLICATION NUMBER: WO PCT/EP91/02409
; FILING DATE: 13-DEC-1991
; APPLICATION NUMBER: EP 90124241.2
; FILING DATE: 14-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 1487-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 7038164000
; TELEFAX: 7038164100
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-941-611-23

Query Match 85.0%; Score 17; DB 10; Length 2894;
Best Local Similarity 100.0%; Pred. No. 2.1e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGVGVEHLEAACNWT 17
|||||
Db 631 MYVGVGVEHLEAACNWT 647

Search completed: November 21, 2003, 22:19:41
Job time : 19.5 secs

; TYPE: PRT					
; ORGANISM: hepatitis C virus					
US-10-128-590-97					
Query Match 85.0%; Score 17; DB 15; Length 363;					
Best Local Similarity 100.0%; Pred. No. 3.5e-10;					
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MYVGGVEHRLAACNWT 17			
DB	248	MYVGGVEHRLAACNWT 264			
RESULT 13					
US-10-187-257-4					
; Sequence 4, Application US/10187257					
; Publication No. US20030138458A1					
; GENERAL INFORMATION:					
; APPLICANT: HOUGHTON, Michael					
; APPLICANT: COATES, Steve					
; APPLICANT: O'HAGAN, Derek					
; TITLE OF INVENTION: HCV E1E2 VACCINE COMPOSITIONS					
; FILE REFERENCE: 2302-17206					
; CURRENT APPLICATION NUMBER: US/10/187,257					
; CURRENT FILING DATE: 2002-06-28					
; NUMBER OF SEQ ID NOS: 5					
; SOFTWARE: PatentIn Ver. 2.0					
; SEQ ID NO 4					
; LENGTH: 637					
; TYPE: PRT					
; ORGANISM: Artificial Sequence					
; FEATURE:					
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region					
; OTHER INFORMATION: amino acid					
US-10-187-257-4					
Query Match 85.0%; Score 17; DB 12; Length 637;					
Best Local Similarity 100.0%; Pred. No. 5.6e-10;					
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MYVGGVEHRLAACNWT 17			
DB	459	MYVGGVEHRLAACNWT 475			
RESULT 14					
US-10-265-083-2					
; Sequence 2, Application US/10265083					
; Publication No. US20030170273A1					
; GENERAL INFORMATION:					
; APPLICANT: O'HAGAN, Derek					
; APPLICANT: VALIANTE, Nicholas					
; TITLE OF INVENTION: ADJUVANT COMPOSITIONS					
; FILE REFERENCE: 2302-18357.30					
; CURRENT APPLICATION NUMBER: US/10/265,083					
; CURRENT FILING DATE: 2002-10-03					
; NUMBER OF SEQ ID NOS: 5					
; SOFTWARE: PatentIn Ver. 2.0					
; SEQ ID NO 2					
; LENGTH: 637					
; TYPE: PRT					
; ORGANISM: Artificial Sequence					
; FEATURE:					
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7					
; OTHER INFORMATION: region					
US-10-265-083-2					
Query Match 85.0%; Score 17; DB 12; Length 637;					
Best Local Similarity 100.0%; Pred. No. 5.6e-10;					
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MYVGGVEHRLAACNWT 17			

; TYPE: PRT					
; ORGANISM: hepatitis C virus					
US-10-128-590-97					
Query Match 85.0%; Score 17; DB 15; Length 363;					
Best Local Similarity 100.0%; Pred. No. 3.5e-10;					
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MYVGGVEHRLAACNWT 17			
DB	248	MYVGGVEHRLAACNWT 264			
RESULT 13					
US-10-187-257-4					
; Sequence 4, Application US/10187257					
; Publication No. US20030138458A1					
; GENERAL INFORMATION:					
; APPLICANT: HOUGHTON, Michael					
; APPLICANT: COATES, Steve					
; APPLICANT: O'HAGAN, Derek					
; TITLE OF INVENTION: HCV E1E2 VACCINE COMPOSITIONS					
; FILE REFERENCE: 2302-17206					
; CURRENT APPLICATION NUMBER: US/10/187,257					
; CURRENT FILING DATE: 2002-06-28					
; NUMBER OF SEQ ID NOS: 5					
; SOFTWARE: PatentIn Ver. 2.0					
; SEQ ID NO 4					
; LENGTH: 637					
; TYPE: PRT					
; ORGANISM: Artificial Sequence					
; FEATURE:					
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region					
; OTHER INFORMATION: amino acid					
US-10-187-257-4					
Query Match 85.0%; Score 17; DB 12; Length 637;					
Best Local Similarity 100.0%; Pred. No. 5.6e-10;					
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MYVGGVEHRLAACNWT 17			
DB	459	MYVGGVEHRLAACNWT 475			
RESULT 14					
US-10-265-083-2					
; Sequence 2, Application US/10265083					
; Publication No. US20030170273A1					
; GENERAL INFORMATION:					
; APPLICANT: O'HAGAN, Derek					
; APPLICANT: VALIANTE, Nicholas					
; TITLE OF INVENTION: ADJUVANT COMPOSITIONS					
; FILE REFERENCE: 2302-18357.30					
; CURRENT APPLICATION NUMBER: US/10/265,083					
; CURRENT FILING DATE: 2002-10-03					
; NUMBER OF SEQ ID NOS: 5					
; SOFTWARE: PatentIn Ver. 2.0					
; SEQ ID NO 2					
; LENGTH: 637					
; TYPE: PRT					
; ORGANISM: Artificial Sequence					
; FEATURE:					
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7					
; OTHER INFORMATION: region					
US-10-265-083-2					
Query Match 85.0%; Score 17; DB 12; Length 637;					
Best Local Similarity 100.0%; Pred. No. 5.6e-10;					
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MYVGGVEHRLAACNWT 17			
DB	248	MYVGGVEHRLAACNWT 264			
RESULT 12					
US-10-128-590-97					
; Sequence 97, Application US/10128590					
; Publication No. US20030108561A1					
; GENERAL INFORMATION:					
; APPLICANT: Innogenetics N.V.					
; TITLE OF INVENTION: Core-glycosylated HCV envelope proteins					
; FILE REFERENCE: 135 PCT					
; CURRENT APPLICATION NUMBER: US/10/128,590					
; CURRENT FILING DATE: 2002-07-22					
; NUMBER OF SEQ ID NOS: 98					
; SOFTWARE: PatentIn version 3.1					
; SEQ ID NO 97					

; TYPE: PRT					
; ORGANISM: hepatitis C virus					
US-10-128-587A-97					
Query Match 85.0%; Score 17; DB 12; Length 363;					
Best Local Similarity 100.0%; Pred. No. 3.5					

```
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-90

Query Match      100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLEAACNWTPE 20.
|||||
Db 1 MYVGGVEHRLEAACNWTPE 20

RESULT 6
US-09-952-572-8
; Sequence 8, Application US/09952572
; Patent No. US20020119495A1
; GENERAL INFORMATION:
; APPLICANT: HAWAII BIOTECHNOLOGY GROUP, Inc.
; APPLICANT: NAKANO, Eileen
; APPLICANT: CLEMENTS, David
; APPLICANT: HUMPHREYS, Tom
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITION OF HEPATITIS C
; FILE REFERENCE: HAWBIO1100
; CURRENT APPLICATION NUMBER: US/09/952,572
; CURRENT FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: US 60/230,927
; PRIOR FILING DATE: 2000-09-13
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 8
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-09-952-572-8

Query Match      85.0%; Score 17; DB 10; Length 250;
Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLEAACNWT 17
|||||
Db 220 MYVGGVEHRLEAACNWT 236

RESULT 7
US-09-407-430-3
; Sequence 3, Application US/09407430
; Patent No. US20020160936A1
; GENERAL INFORMATION:
; APPLICANT: Worman, Howard J.
; APPLICANT: Mamiya, Naoto
; TITLE OF INVENTION: HCV E2 PROTEIN BINDING AGENTS FOR TREATMENT OF
; TITLE OF INVENTION: HEPATITIS C VIRUS INFECTION
; FILE REFERENCE: 0575/54805
; CURRENT APPLICATION NUMBER: US/09/407,430
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 3
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-407-430-3

Query Match      85.0%; Score 17; DB 10; Length 254;
Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLEAACNWT 17
|||||
Db 226 MYVGGVEHRLEAACNWT 242

; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-791-90

Query Match      100.0%; Score 20; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLEAACNWTPE 20.
|||||
Db 1 MYVGGVEHRLEAACNWTPE 20

RESULT 8
US-09-929-955-4
; Sequence 4, Application US/09929955
; Patent No. US20020136740A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.23AUS2
; CURRENT APPLICATION NUMBER: US/09/929,955
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/225,767
; PRIOR FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-09-929-955-4

Query Match      85.0%; Score 17; DB 10; Length 350;
Best Local Similarity 100.0%; Pred. No. 3.4e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLEAACNWT 17
|||||
Db 252 MYVGGVEHRLEAACNWT 268

RESULT 9
US-10-104-966-4
; Sequence 4, Application US/10104966
; Publication NO. US20020155124A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.23AUSC1
; CURRENT APPLICATION NUMBER: US/10/104,966
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus E2 protein sequence
US-10-104-966-4

Query Match      85.0%; Score 17; DB 14; Length 350;
Best Local Similarity 100.0%; Pred. No. 3.4e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLEAACNWT 17
|||||
Db 252 MYVGGVEHRLEAACNWT 268
```

; INFORMATION FOR SEQ ID NO: 90:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 90:
US-09-973-025-90

Query Match 100.0%; Score 20; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLEAACNWTGPE 20
| | | | | | | | | | | | | | | | | | | |
DB 1 MYVGGVEHRLEAACNWTGPE 20

RESULT 2
US-09-899-303-90
; Sequence 90, Application US/09899303
; Publication No. US20030036110A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; BOSMAN, FONS
; DE MARTYNOFF, GUY
; BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: NIXON & VANDERHYE P. C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U. S. A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,303
; FILING DATE: 06-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 90:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 90:
US-09-899-303-90

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLEAACNWTGPE 20
| | | | | | | | | | | | | | | | | | | |

DB 1 MYVGGVEHRLEAACNWTGPE 20

RESULT 3

US-09-995-808-90
; Sequence 90, Application US/09995808
; Publication No. US20030095980A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent in 3.1
; SEQ ID NO 90
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-90

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLEAACNWTGPE 20
| | | | | | | | | | | | | | | | | | | |
DB 1 MYVGGVEHRLEAACNWTGPE 20

RESULT 4

US-09-995-860-90
; Sequence 90, Application US/09995860
; Publication No. US20030118603A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent in 3.1
; SEQ ID NO 90
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-90

Query Match 100.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLEAACNWTGPE 20
| | | | | | | | | | | | | | | | | | | |
DB 1 MYVGGVEHRLEAACNWTGPE 20

RESULT 5

US-09-995-791-90
; Sequence 90, Application US/09995791
; Publication No. US20030147918A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-68
; CURRENT APPLICATION NUMBER: US/09/995,791
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent in 3.1
; SEQ ID NO 90

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:08:28 ; Search time 19.5 Seconds
(without alignments)
187.240 Million cell updates/sec

Title: US-09-973-025-90

Perfect score: 20

Sequence: 1 MYVGVHRLEAACNWTPE 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues.

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications, AA.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	10	US-09-973-025-90
2	20	100.0	20	11	US-09-893-303-90
3	20	100.0	20	11	US-09-995-808-90
4	20	100.0	20	11	US-09-995-860-90
5	20	100.0	20	12	US-09-995-791-90
6	17	85.0	250	10	US-09-952-572-8
7	17	85.0	254	10	US-09-407-430-3
8	17	85.0	350	10	US-09-929-955-4
9	17	85.0	350	14	US-10-104-966-4
10	17	85.0	363	10	US-09-407-430-2
11	17	85.0	363	12	US-10-128-587A-97
12	17	85.0	363	15	US-10-128-590-97
13	17	85.0	637	12	US-10-187-257-4
14	17	85.0	637	12	US-10-265-083-2
15	17	85.0	2894	10	US-09-941-611-23

16	17	85.0	2894	15	US-10-044-995-23	Sequence 23, Appl
17	17	85.0	3011	9	US-09-742-659-4	Sequence 4, Appl
18	17	85.0	3011	9	US-09-916-359-2	Sequence 2, Appl
19	17	85.0	3011	10	US-09-238-076-20	Sequence 20, Appl
20	17	85.0	3011	10	US-09-952-572-9	Sequence 9, Appl
21	17	85.0	3011	10	US-09-929-955-1	Sequence 1, Appl
22	17	85.0	3011	10	US-09-747-419-20	Sequence 20, Appl
23	17	85.0	3011	11	US-09-891-894-3	Sequence 3, Appl
24	17	85.0	3011	11	US-09-995-937-20	Sequence 20, Appl
25	17	85.0	3011	11	US-09-917-563-20	Sequence 20, Appl
26	17	85.0	3011	12	US-10-184-150-3	Sequence 3, Appl
27	17	85.0	3011	14	US-10-104-966-1	Sequence 20, Appl
28	17	85.0	3011	15	US-10-259-275-20	Sequence 6, Appl
29	17	85.0	3011	16	US-10-232-643-6	Sequence 2, Appl
30	17	85.0	3012	10	US-09-238-076-2	Sequence 2, Appl
31	17	85.0	3012	11	US-09-995-937-2	Sequence 2, Appl
32	17	85.0	3012	11	US-09-917-563-2	Sequence 2, Appl
33	14	70.0	176	10	US-09-921-397-81	Sequence 81, Appl
34	12	60.0	32	12	US-10-351-641-1217	Sequence 1217, Ap
35	12	60.0	34	12	US-10-351-641-1715	Sequence 1715, Ap
36	11	55.0	37	12	US-10-351-641-1216	Sequence 1216, Ap
37	11	55.0	39	12	US-10-351-641-1714	Sequence 1714, Ap
38	10	50.0	347	11	US-09-194-949-9	Sequence 9, Appl
39	10	50.0	539	11	US-09-194-949-11	Sequence 11, Appl
40	10	50.0	2940	12	US-10-236-629A-13	Sequence 13, Appl
41	10	50.0	2985	15	US-10-259-275-40	Sequence 40, Appl
42	9	45.0	10	10	US-09-894-018-223	Sequence 223, Appl
43	9	45.0	28	16	US-10-318-200-17	Sequence 17, Appl
44	9	45.0	107	10	US-09-894-018-103	Sequence 103, Appl
45	9	45.0	130	10	US-09-894-018-99	Sequence 99, Appl

ALIGNMENTS

RESULT 1

US-09-973-025-90
; Sequence 90, Application US/09973025
; Publication No. US20020182706A1

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

BOSMAN, FONS

DE MARTYNOFF, GUY

BUYSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/973,025

FILING DATE: 10-Oct-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

THIS PAGE BLANK (USPTO)

APPLICANT: FRAIL, DONALD E.
APPLICANT: DESAI, SURESH M.
APPLICANT: DEVARE, SUSHIL G.
TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/00907
FILING DATE: 19930129
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: FOREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5131.PC.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 333 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-00907-12

Query Match 85.0%; Score 17; DB 5; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGVGVEHRLAACNWT 17
||| ||||| ||||| |||||
Db 280 MYVGVGVEHRLAACNWT 296

Search completed: November 21, 2003, 21:15:18
Job time : 10.6 secs

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: POREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33207
REFERENCE/DOCKET NUMBER: 5763.US.01
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-073-2

Query Match 85.0%; Score 17; DB 3; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.2e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYVGVGVEHRLAACNWT 17
Db 272 MYVGVGVEHRLAACNWT 288

RESULT 13
US-08-453-552-12
Sequence 12, Application US/08453552
Patent No. 5667992
GENERAL INFORMATION:
APPLICANT: CASEY, JAMES M.
APPLICANT: BODE, SUZANNE L.
APPLICANT: ZECK, BILLY J.
APPLICANT: YAMAGUCHI, JULIE
APPLICANT: FRAIL, DONALD E.
APPLICANT: DESAI, SURESH M.
APPLICANT: DEVARE, SUSHIL G.
TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/453,552
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: US/08/144,099
FILING DATE:
APPLICATION NUMBER: US 07/830,024
FILING DATE: 01-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: POREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5131.US.D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9556
TELEFAX: 708-937-6365
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 333 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-453-552-12

Query Match 85.0%; Score 17; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYVGVGVEHRLAACNWT 17
Db 280 MYVGVGVEHRLAACNWT 296

RESULT 14
US-08-710-637-12
Sequence 12, Application US/08710637
Patent No. 5854001
GENERAL INFORMATION:
APPLICANT: CASEY, JAMES M.
APPLICANT: BODE, SUZANNE L.
APPLICANT: ZECK, BILLY J.
APPLICANT: YAMAGUCHI, JULIE
APPLICANT: FRAIL, DONALD E.
APPLICANT: DESAI, SURESH M.
APPLICANT: DEVARE, SUSHIL G.
TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/710,637
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: US/08/144,099
FILING DATE:
APPLICATION NUMBER: US 07/830,024
FILING DATE: 01-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: POREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5131.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9556
TELEFAX: 708-937-6365
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 333 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-710-637-12

Query Match 85.0%; Score 17; DB 2; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYVGVGVEHRLAACNWT 17
Db 280 MYVGVGVEHRLAACNWT 296

RESULT 15
PCT-US93-00907-12
Sequence 12, Application PC/TUS9300907
GENERAL INFORMATION:
APPLICANT: CASEY, JAMES M.
APPLICANT: BODE, SUZANNE L.
APPLICANT: ZECK, BILLY J.
APPLICANT: YAMAGUCHI, JULIE

Query Match 85.0%; Score 17; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLAACNWT 17
|||||
DB 262 MYVGGVEHRLAACNWT 278

RESULT 10
US-08-440-210-15
; Sequence 15, Application US/08440210
; Patent No. 5766845
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions

NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
; ADDRESS: Chiron Corporation

STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA

COUNTRY: USA
ZIP: 94608

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/440,210

FILING DATE: 12-MAY-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/231,368

FILING DATE:

APPLICATION NUMBER: US 07/759,575

FILING DATE: 13-SEP-1991

ATTORNEY/AGENT INFORMATION:

NAME: McClung, Barbara G.

REGISTRATION NUMBER: 33,113

REFERENCE/DOCKET NUMBER: 0205.001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 601-2708

TELEFAX: (510) 655-3542

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 278 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-440-210-15

Query Match 85.0%; Score 17; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLAACNWT 17
|||||
DB 262 MYVGGVEHRLAACNWT 278

RESULT 11
US-09-046-604-15
; Sequence 15, Application US/09046604
; Patent No. 6303292
; GENERAL INFORMATION:

APPLICANT: Weiner, Amy J.

APPLICANT: Houghton, Michael

TITLE OF INVENTION: Immunoreactive Polypeptide Compositions

NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
; ADDRESS: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/046,604
; FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/231,368

FILING DATE:

APPLICATION NUMBER: US 07/759,575

ATTORNEY/AGENT INFORMATION:

NAME: McClung, Barbara G.

REGISTRATION NUMBER: 33,113

REFERENCE/DOCKET NUMBER: 0205.001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 601-2708

TELEFAX: (510) 655-3542

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 278 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-046-604-15

Query Match 85.0%; Score 17; DB 4; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLAACNWT 17
|||||
DB 262 MYVGGVEHRLAACNWT 278

RESULT 12
US-08-478-073-2
; Sequence 2, Application US/08478073
; Patent No. 6020122
; GENERAL INFORMATION:

APPLICANT: Okasinski, Gregory F.

APPLICANT: Schaefer, Verlyn G.

APPLICANT: Suhr, Thomas S.

APPLICANT: Leeniewski, Richard R.

TITLE OF INVENTION: EXPRESSION SYSTEM FOR NON-SECRETOR GENES

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: ABBOTT LABORATORIES

STREET: ONE HUNDRED ABBOTT PARK ROAD

CITY: ABBOTT PARK

STATE: IL

COUNTRY: USA

ZIP: 60064

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/478,073

FILING DATE:


```

RESULT 8
US-08-440-542-15
; Sequence 15, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immuno reactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,542
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424

```

```

; PRIOR APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-542-15

Query Match 85.0%; Score 17; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.le-09;
Matches 17; Conservative 0; Mismatches 0; Indels

QY 1 MYVGGVEHRELEAACNWT 17
Db 262 MYVGGVEHRELEAACNWT 278
|||||
|||||

RESULT 9
US-08-231-368-15
; Sequence 15, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghcon, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-231-368-15

```

;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/07/748,292
;/ FILING DATE: 19910821
;/ CLASSIFICATION: 436
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 07/456,162
;/ FILING DATE: 22-DEC-1989
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 07/610,180
;/ FILING DATE: 07-NOV-1990
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: POREMSKI, PRISCILLA E.
;/ REGISTRATION NUMBER: 33,207
;/ REFERENCE/DOCKET NUMBER: 4767.US.P2
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: 708-937-6365
;/ INFORMATION FOR SEQ ID NO: 1:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 121 amino acids
;/ TYPE: AMINO ACID
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: peptide
;/ US-07-748-292-1

Query Match 85.0%; Score 17; DB 1; Length 121;
Best Local Similarity 100.0%; Pred. No. 5.6e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGVGVEHRLAACNWT 17
Db 32 MYVGVGVEHRLAACNWT 48

RESULT 5
US-08-444-818-56
;/ Sequence 56, Application US/08444818
;/ Patent No. 6150087
;/ GENERAL INFORMATION:
;/ APPLICANT: Chien, David Y.
;/ APPLICANT: Rutter, William J.
;/ TITLE OF INVENTION: NANBV Diagnostics and Vaccines
;/ NUMBER OF SEQUENCES: 777
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Chiron Corporation
;/ STREET: 4560 Horton Street
;/ CITY: Emeryville
;/ STATE: CA
;/ COUNTRY: USA
;/ ZIP: 94608-2916
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: Patent In Release #1.0, Version #1.30
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/444,818
;/ FILING DATE:
;/ CLASSIFICATION: 424
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/403,590
;/ FILING DATE: 14-MAR-1995
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Harbin, Alisa A.
;/ REGISTRATION NUMBER: 33,895
;/ REFERENCE/DOCKET NUMBER: 0110.002
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (508)359-3885
;/ INFORMATION FOR SEQ ID NO: 56:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 141 amino acids

;/ TYPE: amino acid
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: protein
;/ US-08-444-818-56
;/ Query Match 85.0%; Score 17; DB 3; Length 141;
;/ Best Local Similarity 100.0%; Pred. No. 6.4e-10;
;/ Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;/ QY 1 MYVGVGVEHRLAACNWT 17
;/ Db 6 MYVGVGVEHRLAACNWT 22

RESULT 6
US-08-444-818-77
;/ Sequence 77, Application US/08444818
;/ Patent No. 6150087
;/ GENERAL INFORMATION:
;/ APPLICANT: Chien, David Y.
;/ APPLICANT: Rutter, William J.
;/ TITLE OF INVENTION: NANBV Diagnostics and Vaccines
;/ NUMBER OF SEQUENCES: 777
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Chiron Corporation
;/ STREET: 4560 Horton Street
;/ CITY: Emeryville
;/ STATE: CA
;/ COUNTRY: USA
;/ ZIP: 94608-2916
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: Patent In Release #1.0, Version #1.30
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/444,818
;/ FILING DATE:
;/ CLASSIFICATION: 424
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/403,590
;/ FILING DATE: 14-MAR-1995
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Harbin, Alisa A.
;/ REGISTRATION NUMBER: 33,895
;/ REFERENCE/DOCKET NUMBER: 0110.002
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (508)359-3876
;/ INFORMATION FOR SEQ ID NO: 77:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 179 amino acids
;/ TYPE: amino acid
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: protein
;/ US-08-444-818-77

Query Match 85.0%; Score 17; DB 3; Length 179;
Best Local Similarity 100.0%; Pred. No. 7.8e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGVGVEHRLAACNWT 17
Db 121 MYVGVGVEHRLAACNWT 137

RESULT 7
US-08-440-103-15
;/ Sequence 15, Application US/08440103
;/ Patent No. 5670152
;/ GENERAL INFORMATION:
;/ APPLICANT: Weiner, Amy J.
;/ APPLICANT: Houghton, Michael

Db 1 MYVGVVHRLEAACNWTPE 20

RESULT 2

US-08-927-597-90
; Sequence 90, Application US/08927597
; Patent No. 6245503
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 90:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-927-597-90

Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYVGVVHRLEAACNWTPE 20
Db 1 MYVGVVHRLEAACNWTPE 20

RESULT 3

US-08-262-037-47
; Sequence 47, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK

STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/262,037
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/719,819
FILING DATE: 24-June-1991
APPLICATION NUMBER: 07/667,275
FILING DATE: 11-Mar-1991
APPLICATION NUMBER: 07/651,735
FILING DATE: 07-Feb-1991
APPLICATION NUMBER: 07/558,799
FILING DATE: 28-July-1990
APPLICATION NUMBER: 07/510,153
FILING DATE: 16-April-1990
ATTORNEY/AGENT INFORMATION:
NAME: Maria C. H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4043 US3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 39 amino acids
TYPE: Amino acid
STRANDEDNESS:
TOPOLOGY: Unknown
US-08-262-037-47

Query Match 85.0%; Score 17; DB 1; Length 39;
Best Local Similarity 100.0%; Pred. No. 2.2e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYVGVVHRLEAACNWT 17
Db 7 MYVGVVHRLEAACNWT 23

RESULT 4

US-07-748-292-1
; Sequence 1, Application US/07748292
; Patent No. 5308750
; GENERAL INFORMATION:
; APPLICANT: MEHTA, SMRITI U.
; APPLICANT: JOHNSON, JILL E.
; APPLICANT: DAILEY, STEPHEN H.
; APPLICANT: DESAI, SURESH M.
; APPLICANT: DEVARE, SUSHIL G.
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES TO PUTATIVE HCV
; TITLE OF INVENTION: E2/NS1 PROTEINS AND METHODS FOR USING SAME
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D-377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: USA
; ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:47:12 ; Search time 10.6 Seconds
(without alignments)
79.832 Million cell updates/sec

Title: US-09-973-025-90

Perfect score: 20
Sequence: 1 MYVGVHRLAECNWTPE 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2.6/prodata/1/iaa/5A COMB.pep.*
2: /cgn2.6/prodata/1/iaa/5B COMB.pep.*
3: /cgn2.6/prodata/1/iaa/6A COMB.pep.*
4: /cgn2.6/prodata/1/iaa/6B COMB.pep.*
5: /cgn2.6/prodata/1/iaa/PCUS COMB.pep.*
6: /cgn2.6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	3	US-08-612-973-90
2	20	100.0	20	3	US-08-527-597-90
3	17	85.0	39	1	US-08-262-037-47
4	17	85.0	121	1	US-07-748-292-1
5	17	85.0	141	3	US-08-444-818-56
6	17	85.0	179	3	US-08-444-818-77
7	17	85.0	278	1	US-08-440-103-15
8	17	85.0	278	1	US-08-440-542-15
9	17	85.0	278	1	US-08-231-368-15
10	17	85.0	278	1	US-08-440-210-15
11	17	85.0	278	4	US-09-046-604-15
12	17	85.0	305	3	US-08-478-073-2
13	17	85.0	333	1	US-08-453-552-12
14	17	85.0	333	2	US-08-710-637-12
15	17	85.0	333	5	PCT-US93-00907-12
16	17	85.0	337	1	US-08-188-281B-7
17	17	85.0	337	5	PCT-US94-07280-7
18	17	85.0	337	5	PCT-US95-01087-7
19	17	85.0	367	1	US-08-188-281B-9
20	17	85.0	367	1	US-08-453-552-6
21	17	85.0	367	2	US-08-710-637-6
22	17	85.0	367	5	PCT-US93-00907-6
23	17	85.0	367	5	PCT-US94-07280-9
24	17	85.0	367	5	PCT-US95-01087-9
25	17	85.0	377	1	US-08-188-281B-17
26	17	85.0	377	5	PCT-US94-07280-17
27	17	85.0	377	5	PCT-US95-01087-17

28	17	85.0	397	1	US-08-188-281B-11	Sequence 11, Appl
29	17	85.0	397	5	PCT-US94-07280-11	Sequence 11, Appl
30	17	85.0	397	5	PCT-US95-01087-11	Sequence 11, Appl
31	17	85.0	399	1	US-08-453-552-10	Sequence 10, Appl
32	17	85.0	399	2	US-08-710-637-10	Sequence 10, Appl
33	17	85.0	399	5	PCT-US93-00907-10	Sequence 10, Appl
34	17	85.0	402	1	US-08-460-806-4	Sequence 4, Appl
35	17	85.0	402	1	US-08-460-806-5	Sequence 5, Appl
36	17	85.0	402	1	US-08-460-806-13	Sequence 13, Appl
37	17	85.0	402	1	US-08-460-806-15	Sequence 15, Appl
38	17	85.0	402	1	US-08-325-630-4	Sequence 4, Appl
39	17	85.0	402	1	US-08-325-630-5	Sequence 5, Appl
40	17	85.0	402	1	US-08-325-630-13	Sequence 13, Appl
41	17	85.0	402	1	US-08-325-630-15	Sequence 15, Appl
42	17	85.0	403	2	US-08-483-695-39	Sequence 39, Appl
43	17	85.0	403	2	US-07-965-285-39	Sequence 39, Appl
44	17	85.0	403	2	US-08-487-231-39	Sequence 39, Appl
45	17	85.0	403	3	US-09-201-912-39	Sequence 39, Appl

ALIGNMENTS

RESULT 1
US-08-612-973-90
; Sequence 90, Application US/08612973
; Patent No. 6150134
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,973
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 90:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-612-973-90
Query Match 100.0%; Score 20; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MYVGVHRLAECNWTPE 20
|||||

CC region one (HVR1). The HCV E2 protein lacking HVR1 DNA is useful for
 CC producing infectious HCV and chimeric HCV viruses which are useful
 CC for identifying cell lines capable of supporting the replication of
 CC viruses. The infectious HCV and HVR1-chimeric HCV are used in the
 CC production of attenuated or inactivated vaccines which are useful for
 CC treating or preventing HCV in a mammal by immunisation. The host cells
 CC expressing the H77C(HVR1) DNA is useful as an immunogen to stimulate a
 CC protective immune response to HCV. The immunogens are useful for
 CC producing protective antibodies to HCV. The antibodies produced are used
 CC in passive immunoprophylaxis for treatment of diseases caused by HCV in
 CC animals, especially humans. The H77C(HVR1) DNA is also useful in gene
 CC therapy.
 CC Note: The present sequence is not shown in the specification but is
 CC derived from Hepatitis C virus envelope 2 protein lacking hypervariable
 CC region 1 (AAE00442) referred as SEQ ID NO: 2 and shown in figure 1.
 XX
 SQ Sequence 271 AA;

Query Match 85.0%; Score 17; DB 22; Length 271;
 Best Local Similarity 100.0%; Pred. No. 3.7e-10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLAEACNWT 17
 |||||
 Db 241 MYVGGVEHRLAEACNWT 257

RESULT 15
 AAE00446
 ID AAE00446 standard; Protein; 271 AA.
 XX
 AC AAE00446;
 XX
 DT 19-JUN-2001 (first entry)
 XX
 DE HCV E2-634 HVR1-mut containing E1 signal sequence and truncated E2.
 XX
 KW Hepatitis C virus; HCV; hypervariable region one; HVR1; vaccine;
 KW antiviral; gene therapy; envelope 2 protein; E2; immunisation; mutein;
 KW HCV infection; viral replication; passive immunoprophylaxis; mutant.
 XX
 OS Hepatitis C virus.
 OS Synthetic.

XX Key Location/Qualifiers
 FH Peptide 1..20
 FT /label= Signal peptide
 FT /note= "Derived from endoplasmic reticulum E1
 FT signal sequence (364-383 amino acids)"
 FT Protein 21..271
 FT /label= Mature_C_terminal_truncated_E2_protein
 FT Misc-difference 225
 FT /note= "Wild type Leu substituted with His"
 XX
 PN WO200121807-A1.
 XX
 PD 29-MAR-2001.
 XX
 PF 22-SEP-2000; 2000WO-US25987.
 XX
 PR 23-SEP-1999; 99US-0155823.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Forns X, Bukh J, Emerson SU, Purcell RH;
 XX
 DR WPI; 2001-266076/27.
 XX
 PT Novel nucleic acid molecules that encode hepatitis C virus envelope 2
 PT protein lacking all or part of hypervariable region 1 of envelope
 PT protein, useful as vaccine components for treating or preventing HCV
 PT infections

PS Example; Page -; 80pp; English.
 XX
 CC The present sequence is hepatitis C virus (HCV) E2-634 HVR1 construct
 CC containing the endoplasmic reticulum signal sequence of envelope protein
 CC E1 and carboxy-terminal truncated E2 protein lacking the hypervariable
 CC region one (HVR1) with a replacement of leucine for histidine. The HCV
 CC E2 protein lacking HVR1 DNA is useful for producing infectious HCV and
 CC chimeric HCV viruses which are useful for identifying cell lines
 CC capable of supporting the replication of viruses. The infectious HCV and
 CC HVR1-chimeric HCV are used in the production of attenuated or inactivated
 CC vaccines which are useful for treating or preventing HCV in a mammal by
 CC immunisation. The host cells expressing the H77C(HVR1) DNA is useful as
 CC an immunogen to stimulate a protective immune response to HCV. The
 CC immunogens are useful for producing protective antibodies to HCV. The
 CC antibodies produced are used in passive immunoprophylaxis for treatment
 CC of diseases caused by HCV in animals, especially humans. The H77C(HVR1)
 CC DNA is also useful in gene therapy.
 CC Note: The present sequence is not shown in the specification but is
 CC derived from Hepatitis C virus envelope 2 protein lacking hypervariable
 CC region 1 (AAE00442) referred as SEQ ID NO: 2 and shown in figure 1.
 XX
 SQ Sequence 271 AA;

Query Match 85.0%; Score 17; DB 22; Length 271;
 Best Local Similarity 100.0%; Pred. No. 3.7e-10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLAEACNWT 17
 |||||
 Db 241 MYVGGVEHRLAEACNWT 257

Search completed: November 21, 2003, 20:58:07
 Job time : 32.15 secs

PT New immunogenic polypeptide comprising hepatitis C virus E2 polypeptide
 PT useful for treating hepatitis C virus infection and for providing
 PT immune protection against virus infection -
 PS Claim 4; Fig 6; 84pp; English.
 XX
 CC The invention relates to a secreted polypeptide comprising hepatitis C
 CC virus (HCV) E2 polypeptide lacking all or a portion of its membrane
 CC spanning domain so that the E2 polypeptide is capable of secretion into
 CC growth medium when expressed recombinantly in a host cell. The
 CC polypeptide may also lack a portion of its C-terminus. The HCV E2
 CC secreted polypeptide is useful for producing anti-HCV antibodies. A
 CC purified immunogenic polypeptide comprising HCV E2 is useful for treating
 CC HCV infection and for providing immune protection against HCV infection
 CC by administering it to a subject having or at risk of having HCV
 CC infection or in need of protection. The method further comprises
 CC administering an immunomodulatory agent such as interleukin-2 (IL-2),
 CC granulocyte macrophage-colony stimulating factor (GM-CSF) or
 CC gamma-interferon. The polypeptide is useful as a vaccine, and with other
 CC HCV proteins to form a multi-component HCV vaccine for prophylactic or
 CC therapeutic treatment of HCV infection. This sequence represents an N-
 CC and C-terminally truncated HCV delta-delta E2 genotype 1a protein.
 XX
 SQ Sequence 250 AA;

Query Match 85.0%; Score 17; DB 23; Length 250;
 Best Local Similarity 100.0%; Pred. No. 3.5e-10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MYVGGVEHRLAEACNWT 17
 |||||
 DB 220 MYVGGVEHRLAEACNWT 236

RESULT 13
 AAB68043
 ID AAB68043 standard; protein; 254 AA.

AC AAB68043;
 XX
 XX 29-JUN-2001 (first entry)
 DE Amino acid sequence of water soluble variant of envelope E2 protein.
 XX
 XX Eo protein; HCV; envelope protein; E2 protein; HCV infection;
 KW HCV attachment.
 XX
 XX Synthetic.
 OS Hepatitis C virus.
 XX
 PN WO200122984-A1.
 XX
 XX 05-APR-2001.
 PD
 XX
 PF 26-SEP-2000; 2000WO-US26395.
 XX
 PR 29-SEP-1999; 99US-0407430.
 XX
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 PA
 XX Worman HJ, Mamiya N;
 PI
 XX WPI; 2001-273486/28.
 DR
 XX
 XX Treating or preventing hepatitis C virus infection in a subject.
 PT involves administering hepatitis C virus envelope protein E2 binding
 PT agents -
 XX
 PS Claim 5; Fig 8; 46pp; English.

XX The present sequence represents a water soluble variant of a Hepatitis C
 CC virus (HCV) envelope E2 protein. Agents that bind to the HCV E2 protein
 CC (such as the human Eo protein), and so inhibit the attachment of HCV onto

CC cells (especially liver cells), are used to treat HCV infections in
 CC mammals, in particular humans. The specification also describes a method
 CC for identifying a compound which can be used for treating or preventing
 CC HCV in a subject and which can inhibit the attachment of HCV onto cells
 CC by inhibiting the binding of HCV envelope E2 protein to a cellular
 CC protein associated with HCV attachment and entry into cells. The method
 CC comprises incubating the compound, HCV envelope E2 protein or its variant
 CC and a cellular protein capable of specifically binding to the HCV E2
 CC protein under suitable reaction conditions; determining the interactions
 CC between HCV envelope E2 protein and cellular protein in the presence and
 CC absence of the compound; and comparing the interaction to identify a
 CC compound which can inhibit the attachment of HCV onto cells.
 XX

SQ Sequence 254 AA;

Query Match 85.0%; Score 17; DB 22; Length 254;
 Best Local Similarity 100.0%; Pred. No. 3.5e-10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MYVGGVEHRLAEACNWT 17
 |||||
 DB 226 MYVGGVEHRLAEACNWT 242

RESULT 14

AAE00445
 ID AAE00445 standard; Protein; 271 AA.

AC AAE00445;
 XX

DT 19-JUN-2001 (first entry)

XX HCV E2-634 HVR1 construct containing E1 signal sequence and truncated E2.

XX Hepatitis C virus; HCV; hypervariable region one; HVR1; vaccine;
 KW antiviral; gene therapy; envelope 2 protein; E2; immunisation;
 KW HCV infection; viral replication; passive immunoprophylaxis.

XX Hepatitis C virus.
 OS Synthetic.

XX
 FH Key Location/Qualifiers
 FT Peptide 1..20
 FT /label= Signal peptide
 FT /note= "Derived from endoplasmic reticulum E1
 FT signal sequence (364-383 amino acids)"
 FT Protein 21..271
 FT /label= Mature_C_terminal_truncated_E2_protein

XX WO200121807-A1.

XX 29-MAR-2001.

XX 22-SEP-2000; 2000WO-US25987.

XX 23-SEP-1999; 99US-0155823.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Forns X, Bukh J, Emerson SU, Purcell RH;

XX WPI; 2001-266076/27.

XX Novel nucleic acid molecules that encode hepatitis C virus envelope 2
 PT protein lacking all or part of hypervariable region 1 of envelope
 PT protein, useful as vaccine components for treating or preventing HCV
 PT infections -
 XX
 PS Example; Page -; 80pp; English.

XX The present sequence is hepatitis C virus (HCV) E2-634 HVR1 construct
 CC containing the endoplasmic reticulum signal sequence of envelope protein
 CC E1 and carboxy-terminal truncated E2 protein lacking the hypervariable

```

XX PA (MITU ) MITSUBISHI KASEI CORP.
XX PI Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
XX PI Teranishi Y;
XX DR WPI; 1992-417213/51.
XX DR N-PSDB; AAQ32467.
XX PT New hepatitis C virus gene and its encoded protein - used for
XX PT diagnosing and vaccinating against hepatitis C virus infections
XX PS Disclosure; Page 104-05; 305pp; English.
XX CC The sequences given in AAR29543-55 are encoded by various clones of the
XX CC NS1(gp70) region of the Hepatitis C Virus (HCV) gene of the invention.
XX CC These peptides were isolated from the serum of a patient suffering
XX CC from hepatitis C (HC). The RNA sequences encoding these peptides
XX CC were converted into cDNA using transcriptase in the presence of one
XX CC of the primer sequences given in AAQ32547-52. The cDNA sequences
XX CC represent different alleles of the same region of the HCV gene.
XX CC The primer sequences given in AAQ32547-52. The cDNA sequences
XX CC represent different alleles of the same region of the HCV gene.
XX CC The sequences given in AAR29544-52 represent the N19 region of the HCV
XX CC protein and the remaining three peptides represent the MX24 region of
XX CC the protein. The sequences given in AAR29531-32 represent gp70 protein
XX CC analogues which contain the M-gp70 region. M-gp70 is a glycoprotein
XX CC which is located adjacent to the C-terminal of envelope protein,
XX CC M-gp35 and contains a potential transmembrane domain. Therefore, all
XX CC or part of gp70 may be an envelope protein. M-gp70 is relatively
XX CC hydrophobic and has little homology with known HCV amino acid
XX CC sequences. See also AAR29527.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 209 AA;
XX
Query Match 85.0%; Score 17; DB 13; Length 209;
Best Local Similarity 100.0%; Pred. No. 3e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MYVGGVEHRLAACNWT 17
DB 91 MYVGGVEHRLAACNWT 107

```

```

XX PI Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
XX PI Teranishi Y;
XX DR WPI; 1992-417213/51.
XX DR N-PSDB; AAQ32468.
XX PT New hepatitis C virus gene and its encoded protein - used for
XX PT diagnosing and vaccinating against hepatitis C virus infections
XX PS Disclosure; Page 105-06; 305pp; English.
XX CC The sequences given in AAR29543-55 are encoded by various clones of the
XX CC NS1(gp70) region of the Hepatitis C Virus (HCV) gene of the invention.
XX CC These peptides were isolated from the serum of a patient suffering
XX CC from hepatitis C (HC). The RNA sequences encoding these peptides
XX CC were converted into cDNA using transcriptase in the presence of one
XX CC of the primer sequences given in AAQ32547-52. The cDNA sequences
XX CC represent different alleles of the same region of the HCV gene.
XX CC The sequences given in AAR29544-52 represent the N19 region of the HCV
XX CC protein and the remaining three peptides represent the MX24 region of
XX CC the protein. The sequences given in AAR29531-32 represent gp70 protein
XX CC analogues which contain the M-gp70 region. M-gp70 is a glycoprotein
XX CC which is located adjacent to the C-terminal of envelope protein,
XX CC M-gp35 and contains a potential transmembrane domain. Therefore, all
XX CC or part of gp70 may be an envelope protein. M-gp70 is relatively
XX CC hydrophobic and has little homology with known HCV amino acid
XX CC sequences. See also AAR29527.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 209 AA;
XX
Query Match 85.0%; Score 17; DB 13; Length 209;
Best Local Similarity 100.0%; Pred. No. 3e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MYVGGVEHRLAACNWT 17
DB 91 MYVGGVEHRLAACNWT 107

```

RESULT 12

```

AAU79220
ID AAU79220 standard; Protein; 250 AA.
XX AC AAU79220;
XX DT 15-JUL-2002 (first entry)
XX DE Hepatitis C Virus (HCV) delta-delta E2 genotype 1a protein.
XX KW Hepatitis C Virus E2; HCV E2; virucide; hepatotropic; IL-2;
XX KW antiinflammatory; HCV infection; interleukin-2; gamma-interferon;
XX KW granulocyte macrophage-colony stimulating factor; GM-CSF;
XX KW delta-delta E2 genotype 1a.
XX OS Hepatitis C Virus.
XX PN WO200222155-A1.
XX PD 21-MAR-2002.
XX PF 13-SEP-2001; 2001WO-US28767.
XX PR 13-SEP-2000; 2000US-230927P.
XX PA (HAWA-) HAWAII BIOTECHNOLOGY GROUP INC.
XX PI Nakano ET, Clements DE, Humphreys T;
XX PR WPI; 2002-383102/41.
XX DR N-PSDB; ABK49386.
XX

```

PD 13-SEP-2000.
XX
PF 16-MAR-1990; 2000EP-0109602.
XX
PR 17-MAR-1989; 89US-0325338.
PR 20-APR-1989; 89US-0341334.
PR 18-MAY-1989; 89US-0355002.
XX 16-MAR-1990; 90EP-0302866.
XX
PA (CHIR) CHIRON CORP.
XX
XX
PI Houghton M, Choo Q, Kuo G;
XX
XX WPI; 2000-566891/53.
DR N-PSDB; AAA75285.
XX
XX

PT Novel composition comprising a hepatitis C virus antisense
PT polynucleotide which is complementary to or corresponds to a sense
PT strand of the virus genome, and selectively hybridizes to it -
XX
XX Example; Fig 5; 75pp; English.
XX

CC The specification describes a pharmaceutical composition which
CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
CC HCV is characterized by a positive stranded RNA genome which has
CC 40% homology at the polypeptide level to a HCV polypeptide. The
CC antisense polynucleotide binds to cellular polynucleotides which
CC enhance and/or are required for viral infectivity, replicative
CC ability or chronicity. The antisense polynucleotides may also be
CC designed to bind with high specificity, to be of increased stability,
CC to be stable and to have low toxicity. The composition also comprises
CC an agent which causes viral RNA to be inactive. The composition
CC is used for preventing HCV replication in a system. The present
CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
CC course of the invention.
XX
XX

SQ Sequence 179 AA;

Query Match 85.0%; Score 17; DB 21; Length 179;
Best Local Similarity 100.0%; Pred. No. 2.6e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLAECNWT 17
|||||
Db 121 MYVGGVEHRLAECNWT 137

RESULT 9
AAR29553
ID AAR29553 standard; Protein; 209 AA.
XX
XX

AC AAR29553;

DT 25-MAR-2003 (updated)
DT 26-APR-1993 (first entry)

DE HCV NS1(gp70) peptide MX24-4.

XX Clone; polypeptide; NS1(gp70); Hepatitis C; Virus; HCV; serum; HC;
KW transcriptase; cDNA; primer; allele; N27; N19; MX24; gp70; M-gp70;
KW transmembrane; envelope; glycoprotein; analogue.
XX

OS Hepatitis C virus.

XX
XX PN EP518313-A2.

XX
XX PD 16-DEC-1992.

XX
XX PF 11-JUN-1992; 92EP-0109812.

XX
XX PR 11-JUN-1991; 91JP-0139268.

PR 12-JUL-1991; 91JP-0172794.
PR 07-OCT-1991; 91JP-0287008.
PR

PR 16-DEC-1991; 91JP-0332329.
PR 20-APR-1992; 92JP-0099957.
XX
XX PA (MITU) MITSUBISHI KASEI CORP.
XX
XX Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
PI Teranishi Y;
XX
XX WPI; 1992-417213/51.
DR N-PSDB; AAQ32466.
XX
XX
PT New hepatitis C virus gene and its encoded protein - used for
PT diagnosing and vaccinating against hepatitis C virus infections
XX
XX PS Disclosure; Page 102-03; 305pp; English.
XX

CC The sequences given in AAR29543-55 are encoded by various clones of the
CC NS1(gp70) region of the Hepatitis C Virus (HCV) gene of the invention.
CC These peptides were isolated from the serum of a patient suffering
CC from hepatitis C (HC). The RNA sequences encoding these peptides
CC were converted into cDNA using transcriptase in the presence of one
CC of the primer sequences given in AAQ32547-52. The cDNA sequences
CC represent different alleles of the same region of the HCV gene.
CC The sequences given in AAR29544-52 represent the N19 region of the HCV
CC protein and the remaining three peptides represent the MX24 region of
CC the protein. The sequences given in AAR29531-32 represent gp70 protein
CC analogues which contain the M-gp70 region. M-gp70 is a glycoprotein
CC which is located adjacent to the C-terminal of envelope protein,
CC M-gp15 and contains a potential transmembrane domain. Therefore, all
CC or part of gp70 may be an envelope protein. M-gp70 is relatively
CC hydrophobic and has little homology with known HCV amino acid
CC sequences. See also AAR29527.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
XX

SQ Sequence 209 AA;

Query Match 85.0%; Score 17; DB 13; Length 209;
Best Local Similarity 100.0%; Pred. No. 3e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVEHRLAECNWT 17
|||||
Db 91 MYVGGVEHRLAECNWT 107

RESULT 10
AAR29554
ID AAR29554 standard; Protein; 209 AA.
XX
XX

AC AAR29554;

DT 25-MAR-2003 (updated)
DT 26-APR-1993 (first entry)

XX HCV NS1(gp70) peptide MX24-5.

XX Clone; polypeptide; NS1(gp70); Hepatitis C; Virus; HCV; serum; HC;
KW transcriptase; cDNA; primer; allele; N27; N19; MX24; gp70; M-gp70;
KW transmembrane; envelope; glycoprotein; analogue.
XX

OS Hepatitis C virus.

XX
XX PN EP518313-A2.

XX
XX PD 16-DEC-1992.

XX
XX PF 11-JUN-1992; 92EP-0109812.

XX
XX PR 11-JUN-1991; 91JP-0139268.

PR 12-JUL-1991; 91JP-0172794.
PR 07-OCT-1991; 91JP-0287008.
PR 16-DEC-1991; 91JP-0332329.
PR 20-APR-1992; 92JP-0099957.

DE Protein encoded by novel hepatitis C virus cDNA clone 12f.
 XX
 KW Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
 KW viral infectivity; viral replication.
 XX
 OS Hepatitis C virus.
 XX
 PN EP1034785-A2.
 XX
 PD 13-SEP-2000.
 XX
 PF 16-MAR-1990; 2000EP-0109602.
 XX
 PR 17-MAR-1989; 89US-0325338.
 PR 20-APR-1989; 89US-0341334.
 PR 18-MAY-1989; 89US-0355002.
 PR 16-MAR-1990; 90EP-0302866.
 XX
 XX (CHIR) CHIRON CORP.
 XX
 XX Houghton M, Choo Q, Kuo G;
 PI
 XX WPI; 2000-566891/53.
 DR N-PSDB; AAA75281.
 XX
 XX Novel composition comprising a hepatitis C virus antisense
 PT polynucleotide which is complementary to or corresponds to a sense
 PT strand of the virus genome, and selectively hybridises to it -
 XX
 XX Example; Fig 1; 75pp; English.
 XX
 CC The specification describes a pharmaceutical composition which
 CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
 CC HCV is characterized by a positive stranded RNA genome which has
 CC 40% homology at the polypeptide level to a HCV polyprotein. The
 CC antisense polynucleotide binds to cellular polynucleotides which
 CC enhance and/or are required for viral infectivity, replicative
 CC ability or chronicity. The antisense polynucleotides may also be
 CC designed to bind with high specificity, to be of increased stability,
 CC to be stable and to have low toxicity. The composition also comprises
 CC an agent which causes viral RNA to be inactive. The composition
 CC is used for preventing HCV replication in a system. The present
 CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
 CC course of the invention.
 XX
 SQ Sequence 141 AA;

RESULT 6
 AAP92042
 ID AAP92042 standard; protein; 141 AA.
 XX
 AC AAP92042;
 XX
 DT 25-MAR-2003 (updated)
 DT 02-MAR-1990 (first entry)
 XX
 XX Sequence encoded in the hepatitis C virus (HCV) cDNA insert in clone 12f.
 DE
 XX Hepatitis C virus (HCV); non-A, non-B hepatitis (HAMBH)
 KW
 XX Hepatitis C virus.
 OS
 XX
 XX
 FH Key Location/Qualifiers
 FT Region 92..141
 FT /*tag = a

PN EP318216-A.
 XX
 PD 31-MAY-1989.
 XX
 PF 18-NOV-1988; 88EP-0310922.
 XX
 PR 18-NOV-1987; 87US-0122714.
 PR 30-DEC-1987; 87US-0139886.
 PR 26-FEB-1988; 88US-0161072.
 PR 06-MAY-1988; 88US-0191263.
 PR 26-OCT-1988; 88US-0263584.
 PR 14-NOV-1988; 88US-0271450.
 XX
 XX (CHIR) CHIRON CORP.
 PA
 XX Houghton M, Choo Q, Kuo G;
 PI
 XX WPI; 1989-159274/22.
 DR N-PSDB; AAN92098.
 XX
 XX Purified hepatitis C virus
 PT - and associated nucleic acids and polypeptide(s)
 XX
 XX Claim 13; Figure 27; 139pp; English.
 PS
 XX It is the sequence encoded in the hepatitis C virus (HCV) cDNA insert in
 CC clone 12f. Tag a = the region of overlap with the HCV antigen encoded in
 CC clone 141. It is antigenic and could be used in immunoassay reagents
 CC and vaccines and to generate antibodies useful in diagnosis and passive
 CC immunotherapy for HCV infection/non-A, non-B hepatitis.
 CC (Updated on 25-MAR-2003 to correct PR field.)
 CC (Updated on 25-MAR-2003 to correct PI field.)
 XX
 SQ Sequence 141 AA;

Query Match 85.0%; Score 17; DB 10; Length 141;
 Best Local Similarity 100.0%; Pred. No. 2.1e-10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVHRLEAACNWT 17
 |||||
 DB 6 MYVGGVHRLEAACNWT 22

RESULT 7
 AAB18525
 ID AAB18525 standard; protein; 141 AA.
 XX
 AC AAB18525;
 XX
 DT 15-JAN-2001 (first entry)
 XX

Query Match 85.0%; Score 17; DB 21; Length 141;
 Best Local Similarity 100.0%; Pred. No. 2.1e-10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGGVHRLEAACNWT 17
 |||||
 DB 6 MYVGGVHRLEAACNWT 22

RESULT 8
 AAB18529
 ID AAB18529 standard; protein; 179 AA.
 XX
 AC AAB18529;
 XX
 DT 15-JAN-2001 (first entry)
 XX
 DE Protein encoded by a novel hepatitis C virus cDNA clone 13i.
 XX
 KW Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
 KW viral infectivity; viral replication.
 XX
 OS Hepatitis C virus.
 OS
 XX EP1034785-A2.
 XX

CC polynucleotides are referred to as a Savine. The synthetic polypeptide is
 CC useful for modulating immune responses preferably directed against a
 CC pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix,
 CC colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone
 CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.
 CC Compositions comprising the polypeptide may be used in the treatment or
 CC prophylaxis against viral (such as infections caused by HIV (human
 CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
 CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
 CC (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,
 CC Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic
 CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
 CC Trypanosoma, Toxoplasma and Giardia) infections. The present
 CC sequence is a peptide derived from a parent protein used to
 CC construct a savine of the invention.

SQ Sequence 30 AA;

Query Match 85.0%; Score 17; DB 23; Length 30;
 Best Local Similarity 100.0%; Pred. No. 5.9e-11;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGVGVEHRLAEACNWT 17
 |||||
 DB 3 MYVGVGVEHRLAEACNWT 19

RESULT 4

ID AAR32418 standard; peptide; 121 AA.

XX AAR32418;

XX 25-MAR-2003 (updated)

DT 03-JUL-1993 (first entry)

XX Sequence of immunogenic domain of E2/NS1 region of hepatitis C virus
 DE (HCV) genome encompassing AAs 600-720.

XX Immunogenic peptide; hepatitis C virus; immunogenic domain;
 KW monoclonal antibody; diagnosis; detection; therapy.

XX Synthetic.

OS WO9304205-A1.

PN 04-MAR-1993.

PD 21-AUG-1992; 92WO-US07189.

PF 21-AUG-1991; 91US-0748292.

PR (ABBO) ABBOTT LAB.

XX Dailey SH, Desai SM, Devare SG, Johnson JE, Mehta SU;

XX WPI; 1993-094030/11.

XX Monoclonal antibodies specific for hepatitis C virus E2-NS1
 PT antigen - useful for diagnosis and evaluation of HCV infections
 PT and in differentiation studies

XX Example; Page 29; 48pp; English.

XX Immunogenic domains of E2/NS1 region of HCV genome encompassing AAs
 CC 600-720 were mapped with PFSCAN analysis. FAB dimers of IgG
 CC purified from sera of individuals seropositive for antibodies to HCV
 CC proteins were used as the primary antibody for the serological
 CC analysis. Based on the reactivity these sera in EIA, four AA
 CC sequences (AAR32419, R32420, AAR33182, R33183) were identified as
 CC immunogenic domains. Each of these four sequences and an additional
 CC sequence, which was the combination of the two most immunogenic
 CC sequences (AAR33184) were synthesised. Based on the EIA reactivity of

CC a panel of HCV positive sera, peptide AAR33184 was chosen as the
 CC immunogen for the generation of monoclonal antibodies to HCV NS1.
 CC (Updated on 25-MAR-2003 to correct PN field.)

SQ Sequence 121 AA;

Query Match 85.0%; Score 17; DB 14; Length 121;
 Best Local Similarity 100.0%; Pred. No. 1.9e-10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGVGVEHRLAEACNWT 17
 |||||
 DB 32 MYVGVGVEHRLAEACNWT 48

RESULT 5

ID AAP90159 standard; protein; 141 AA.

XX AAP90159;

XX 25-MAR-2003 (updated)

DT 01-NOV-1989 (first entry)

XX Sequence of hepatitis C virus CDNA insert in clone 12f.

XX Hepatitis C virus; clone 14i; clone 12f; probe; vaccine.

XX Pan troglodytes.

FT Key Location/Qualifiers
 FT Region 92..141

XX GB2212511-A.

XX 26-JUL-1989.

XX 18-NOV-1988; 88GB-0027024.

XX 18-NOV-1987; 87US-0122714.

PR 30-DEC-1987; 87US-0139886.

PR 26-FEB-1988; 88US-0161072.

PR 26-OCT-1988; 88US-0263584.

XX (CHIR) CHIRON CORPORATION.

XX Houghton M, Choo QL, Kuo G;

DR WPI; 1989-215054/30.

DR N-PSDB; AAN90328.

XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
 PT polypeptide(s) and antibodies for diagnosis, prevention and
 PT treatment of infection.

XX Disclosure; fig 27; 235pp; English.

XX The sequence is the peptide encoded by the hepatitis C virus
 CC (HCV) CDNA insert in clone 12f (see AAN90328). The polypeptides
 CC are used to diagnose HCV-induced NANBH, to raise antibodies for
 CC immunosay or treatment, or to produce vaccines.

CC The region shown overlaps with clone 14i.

CC (Updated on 25-MAR-2003 to correct PR field.)

XX Sequence 141 AA;

Query Match 85.0%; Score 17; DB 10; Length 141;
 Best Local Similarity 100.0%; Pred. No. 2.1e-10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVGVGVEHRLAEACNWT 17
 |||||
 DB 6 MYVGVGVEHRLAEACNWT 22

```

PT assays of HCV
XX
PS Example 7; Page 67; 146pp; English.
XX
CC AAR90965-R90980 and AAR90995-R91015 represent synthetic hepatitis C
CC virus (HCV) E1 and E2 peptides used in competition studies. This
CC sequence represents a synthetic E2 peptide, and corresponds to residues
CC 631-650 of the E2 protein sequence. These sequences are useful for in
CC vitro monitoring of HCV disease, or prognosis of the response to
CC interferon treatment of patients suffering from HCV infection. These
CC sequences compete with the proteins produced by AAT12704-T12709 and
CC AAT12961-T12974, which are included in vectors for the production of
CC recombinant E1, E2, and E1/E2. The recombinant proteins can then be
CC isolated and purified by carrying out a disulphide bond cleavage, or a
CC reduction step with a disulphide bond cleavage agent, after lysis of
CC recombinant host cells. The constructs containing the purified HCV
CC envelope proteins can be used for vaccinating humans against HCV, for in
CC vitro detection of HCV antibodies in a sample, and in a serotyping assay
CC for detecting one or more serological types of HCV present in a
CC biological sample. The constructs can also be immobilised on a solid
CC substrate and incorporated into a reversed phase hybridisation assay for
CC determining the presence or the genotype of HCV. The new purification
CC method preserves the conformation of the recombinantly expressed E1, E2
CC and E1/E2, and eliminates contaminating proteins. Antigens isolated
CC using this method are more reactive with human sera than those isolated
CC by known techniques.
XX
SQ Sequence 20 AA;
Query Match 100.0%; Score 20; DB 17; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MYVGGVEHRLAEACNWTGPE 20
| | | | | | | | | | | | | | | |
DB 1 MYVGGVEHRLAEACNWTGPE 20
| | | | | | | | | | | | | | | |

RESULT 2
AAO18719
ID AAO18719 standard; Peptide; 20 AA.
XX
AC AAO18719;
XX
XX 24-OCT-2002 (first entry)
XX
DE Hepatitis C virus E2 protein derived peptide E2-31.
XX
XX Hepatitis C virus; HCV; E1 protein; E2 protein; infection; virucide;
KW immunostimulant; vaccine.
XX
OS Hepatitis C virus.
XX
PN WO200255548-A2.
XX
PD 18-JUL-2002.
XX
XX 11-JAN-2002; 2002WO-EP00219.
XX
XX 11-JAN-2001; 2001US-260699P.
XX
XX 30-AUG-2001; 2001US-315768P.
XX
XX (INNO-) INNOGENETICS NV.
XX
XX Maertens G, Bosman F, Buyse M;
XX
XX WPI; 2002-599657/64.
XX
XX New therapeutic vaccine compositions comprising at least one purified
XX recombinant hepatitis C virus (HCV) single or specific oligomeric
XX recombinant envelope protein E1 or E2, useful for immunizing humans
XX from HCV infection
XX
PS The present invention relates to new therapeutic vaccine compositions for
XX inducing hepatitis C virus (HCV)-specific antibodies, comprising a
XX composition containing at least one purified recombinant HCV single or
XX specific oligomeric recombinant envelope proteins selected from an E1 and
XX an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are
XX useful for inducing HCV-specific antibodies or for immunising humans
XX against HCV. The recombinant HCV E1 and/or E2 proteins are useful as
XX vaccines or therapeutics, in HCV screening and confirmatory antibody
XX tests, for raising antibodies, in the preparation of medicament, and for
XX in vitro monitoring of HCV disease or prognosing the response to
XX treatment of patients suffering from HCV infection. The present sequence
XX is a peptide derived from the proteins of the invention.
XX
SQ Sequence 20 AA;
Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MYVGGVEHRLAEACNWTGPE 20
| | | | | | | | | | | | | | | |
DB 1 MYVGGVEHRLAEACNWTGPE 20
| | | | | | | | | | | | | | | |

RESULT 3
AAU84640
ID AAU84640 standard; Peptide; 30 AA.
XX
AC AAU84640;
XX
XX 08-MAY-2002 (first entry)
XX
XX HCV HepC1a segment 43.
XX
XX Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
XX viral infection; human immunodeficiency virus; melanoma;
XX bacterial infection; Salmonella; Legionella; parasitic infection;
XX Trypanosoma; Toxoplasma; Giardia.
XX
XX Hepatitis C virus.
XX
XX WO2001019197-A1.
XX
XX 29-NOV-2001.
XX
XX 25-MAY-2001; 2001WO-AU00622.
XX
XX 26-MAY-2000; 2000AU-0007761.
XX
XX (AUSU ) UNIV AUSTRALIAN NAT.
XX
XX Thomson SA, Ramshaw IA;
XX
XX WPI; 2002-147575/19.
XX
XX N-PSDB; ABK36478.
XX
XX New synthetic polypeptides having several different segments of at
XX least one parent polypeptide linked together differently compared to
XX the linkage in the parent polypeptide, for inducing immune response
XX against a pathogen or cancer
XX
XX Example 2; Fig 26; 364pp; English.
XX
XX The invention relates to a new synthetic polypeptide (I) comprising
XX several different segments of at least one parent polypeptide linked
XX together in a different relationship relative to their linkage in the
XX parent polypeptide to impede, abrogate or otherwise alter at least one
XX function associated with the parent polypeptide and for inducing an
XX immune response against a pathogen or cancer. Also included are a
XX synthetic polynucleotide encoding and a computer system for
XX designing the synthetic polypeptides. The synthetic polypeptides and

```



```

OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94309140; PubMed=7519526;
RA Kato N., Ootsuyama Y., Sekiya H., Ohkoshi S., Nakazawa T.,
RA Hijikata M., Shimotohno K.;
RT "Genetic drift in hypervariable region 1 of the viral genome in
RT persistent hepatitis C virus infection.";
RL J. Virol. 68:4776-4784 (1994).
DR ENBL; D26409; EAA05425.1; .
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR Coats; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 133
SQ SEQUENCE 133 AA; 14768 MW; FC9475DBD38C1657 CRC64;

Query Match 70.0%; Score 14; DB 12; Length 133;
Best Local Similarity 100.0%; Pred. No. 4.2e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 TIFKIRMYGGVEH 20
Db 115 TIFKIRMYGGVEH 128

```

Search completed: November 21, 2003, 21:08:21
Job time : 24.45 secs

RESULT 13

Q91ZA4 PRELIMINARY; PRT; 3033 AA.
 AC Q91ZA4;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD2a-4;
 RA Itakura J., Nagayama K., Enomoto N., Kurosaki M., Watanabe H.,
 RA Sato C.;
 RT "Full length cDNA sequence of HCV genotype 2a, strain MD2a-4.";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL; AF238483; AAF59942.1; -;
 DR HSSP; P27958; 1A1V.
 DR InterPro; IPR000345; CytC_heme_bind.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_core.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01538; HCV NS1; 1.
 DR Pfam; PF01537; HCV NS2; 1.
 DR Pfam; PF01536; HCV NS3; 1.
 DR Pfam; PF01535; HCV NS4a; 1.
 DR Pfam; PF01534; HCV NS4b; 1.
 DR Pfam; PF01533; HCV NS5a; 1.
 DR Pfam; PF01532; HCV RdRp; 1.
 DR Pfam; PF01531; HCV NS1; 1.
 DR Pfam; PF01530; HCV NS2; 1.
 DR Pfam; PF01529; HCV NS3; 1.
 DR Pfam; PF01528; HCV NS4a; 1.
 DR Pfam; PF01527; HCV NS4b; 1.
 DR Pfam; PF01526; HCV NS5a; 1.
 DR Pfam; PF01525; HCV RdRp; 1.
 DR Pfam; PF01524; HCV NS1; 1.
 DR Pfam; PF01523; HCV NS2; 1.
 DR Pfam; PF01522; HCV NS3; 1.
 DR Pfam; PF01521; HCV NS4a; 1.
 DR Pfam; PF01520; HCV NS4b; 1.
 DR Pfam; PF01519; HCV NS5a; 1.
 DR Pfam; PF01518; HCV RdRp; 1.
 DR Pfam; PF01517; HCV NS1; 1.
 DR Pfam; PF01516; HCV NS2; 1.
 DR Pfam; PF01515; HCV NS3; 1.
 DR Pfam; PF01514; HCV NS4a; 1.
 DR Pfam; PF01513; HCV NS4b; 1.
 DR Pfam; PF01512; HCV NS5a; 1.
 DR Pfam; PF01511; HCV RdRp; 1.
 DR Pfam; PF01510; HCV NS1; 1.
 DR Pfam; PF01509; HCV NS2; 1.
 DR Pfam; PF01508; HCV NS3; 1.
 DR Pfam; PF01507; HCV NS4a; 1.
 DR Pfam; PF01506; HCV NS4b; 1.
 DR Pfam; PF01505; HCV NS5a; 1.
 DR Pfam; PF01504; HCV RdRp; 1.
 DR Pfam; PF01503; HCV NS1; 1.
 DR Pfam; PF01502; HCV NS2; 1.
 DR Pfam; PF01501; HCV NS3; 1.
 DR Pfam; PF01500; HCV NS4a; 1.
 DR Pfam; PF01499; HCV NS4b; 1.
 DR Pfam; PF01498; HCV NS5a; 1.
 DR Pfam; PF01497; HCV RdRp; 1.
 DR Pfam; PF01496; HCV NS1; 1.
 DR Pfam; PF01495; HCV NS2; 1.
 DR Pfam; PF01494; HCV NS3; 1.
 DR Pfam; PF01493; HCV NS4a; 1.
 DR Pfam; PF01492; HCV NS4b; 1.
 DR Pfam; PF01491; HCV NS5a; 1.
 DR Pfam; PF01490; HCV RdRp; 1.
 DR Pfam; PF01489; HCV NS1; 1.
 DR Pfam; PF01488; HCV NS2; 1.
 DR Pfam; PF01487; HCV NS3; 1.
 DR Pfam; PF01486; HCV NS4a; 1.
 DR Pfam; PF01485; HCV NS4b; 1.
 DR Pfam; PF01484; HCV NS5a; 1.
 DR Pfam; PF01483; HCV RdRp; 1.
 DR Pfam; PF01482; HCV NS1; 1.
 DR Pfam; PF01481; HCV NS2; 1.
 DR Pfam; PF01480; HCV NS3; 1.
 DR Pfam; PF01479; HCV NS4a; 1.
 DR Pfam; PF01478; HCV NS4b; 1.
 DR Pfam; PF01477; HCV NS5a; 1.
 DR Pfam; PF01476; HCV RdRp; 1.
 DR Pfam; PF01475; HCV NS1; 1.
 DR Pfam; PF01474; HCV NS2; 1.
 DR Pfam; PF01473; HCV NS3; 1.
 DR Pfam; PF01472; HCV NS4a; 1.
 DR Pfam; PF01471; HCV NS4b; 1.
 DR Pfam; PF01470; HCV NS5a; 1.
 DR Pfam; PF01469; HCV RdRp; 1.
 DR Pfam; PF01468; HCV NS1; 1.
 DR Pfam; PF01467; HCV NS2; 1.
 DR Pfam; PF01466; HCV NS3; 1.
 DR Pfam; PF01465; HCV NS4a; 1.
 DR Pfam; PF01464; HCV NS4b; 1.
 DR Pfam; PF01463; HCV NS5a; 1.
 DR Pfam; PF01462; HCV RdRp; 1.
 DR Pfam; PF01461; HCV NS1; 1.
 DR Pfam; PF01460; HCV NS2; 1.
 DR Pfam; PF01459; HCV NS3; 1.
 DR Pfam; PF01458; HCV NS4a; 1.
 DR Pfam; PF01457; HCV NS4b; 1.
 DR Pfam; PF01456; HCV NS5a; 1.
 DR Pfam; PF01455; HCV RdRp; 1.
 DR Pfam; PF01454; HCV NS1; 1.
 DR Pfam; PF01453; HCV NS2; 1.
 DR Pfam; PF01452; HCV NS3; 1.
 DR Pfam; PF01451; HCV NS4a; 1.
 DR Pfam; PF01450; HCV NS4b; 1.
 DR Pfam; PF01449; HCV NS5a; 1.
 DR Pfam; PF01448; HCV RdRp; 1.
 DR Pfam; PF01447; HCV NS1; 1.
 DR Pfam; PF01446; HCV NS2; 1.
 DR Pfam; PF01445; HCV NS3; 1.
 DR Pfam; PF01444; HCV NS4a; 1.
 DR Pfam; PF01443; HCV NS4b; 1.
 DR Pfam; PF01442; HCV NS5a; 1.
 DR Pfam; PF01441; HCV RdRp; 1.
 DR Pfam; PF01440; HCV NS1; 1.
 DR Pfam; PF01439; HCV NS2; 1.
 DR Pfam; PF01438; HCV NS3; 1.
 DR Pfam; PF01437; HCV NS4a; 1.
 DR Pfam; PF01436; HCV NS4b; 1.
 DR Pfam; PF01435; HCV NS5a; 1.
 DR Pfam; PF01434; HCV RdRp; 1.
 DR Pfam; PF01433; HCV NS1; 1.
 DR Pfam; PF01432; HCV NS2; 1.
 DR Pfam; PF01431; HCV NS3; 1.
 DR Pfam; PF01430; HCV NS4a; 1.
 DR Pfam; PF01429; HCV NS4b; 1.
 DR Pfam; PF01428; HCV NS5a; 1.
 DR Pfam; PF01427; HCV RdRp; 1.
 DR Pfam; PF01426; HCV NS1; 1.
 DR Pfam; PF01425; HCV NS2; 1.
 DR Pfam; PF01424; HCV NS3; 1.
 DR Pfam; PF01423; HCV NS4a; 1.
 DR Pfam; PF01422; HCV NS4b; 1.
 DR Pfam; PF01421; HCV NS5a; 1.
 DR Pfam; PF01420; HCV RdRp; 1.
 DR Pfam; PF01419; HCV NS1; 1.
 DR Pfam; PF01418; HCV NS2; 1.
 DR Pfam; PF01417; HCV NS3; 1.
 DR Pfam; PF01416; HCV NS4a; 1.
 DR Pfam; PF01415; HCV NS4b; 1.
 DR Pfam; PF01414; HCV NS5a; 1.
 DR Pfam; PF01413; HCV RdRp; 1.
 DR Pfam; PF01412; HCV NS1; 1.
 DR Pfam; PF01411; HCV NS2; 1.
 DR Pfam; PF01410; HCV NS3; 1.
 DR Pfam; PF01409; HCV NS4a; 1.
 DR Pfam; PF01408; HCV NS4b; 1.
 DR Pfam; PF01407; HCV NS5a; 1.
 DR Pfam; PF01406; HCV RdRp; 1.
 DR Pfam; PF01405; HCV NS1; 1.
 DR Pfam; PF01404; HCV NS2; 1.
 DR Pfam; PF01403; HCV NS3; 1.
 DR Pfam; PF01402; HCV NS4a; 1.
 DR Pfam; PF01401; HCV NS4b; 1.
 DR Pfam; PF01400; HCV NS5a; 1.
 DR Pfam; PF01399; HCV RdRp; 1.
 DR Pfam; PF01398; HCV NS1; 1.
 DR Pfam; PF01397; HCV NS2; 1.
 DR Pfam; PF01396; HCV NS3; 1.
 DR Pfam; PF01395; HCV NS4a; 1.
 DR Pfam; PF01394; HCV NS4b; 1.
 DR Pfam; PF01393; HCV NS5a; 1.
 DR Pfam; PF01392; HCV RdRp; 1.
 DR Pfam; PF01391; HCV NS1; 1.
 DR Pfam; PF01390; HCV NS2; 1.
 DR Pfam; PF01389; HCV NS3; 1.
 DR Pfam; PF01388; HCV NS4a; 1.
 DR Pfam; PF01387; HCV NS4b; 1.
 DR Pfam; PF01386; HCV NS5a; 1.
 DR Pfam; PF01385; HCV RdRp; 1.
 DR Pfam; PF01384; HCV NS1; 1.
 DR Pfam; PF01383; HCV NS2; 1.
 DR Pfam; PF01382; HCV NS3; 1.
 DR Pfam; PF01381; HCV NS4a; 1.
 DR Pfam; PF01380; HCV NS4b; 1.
 DR Pfam; PF01379; HCV NS5a; 1.
 DR Pfam; PF01378; HCV RdRp; 1.
 DR Pfam; PF01377; HCV NS1; 1.
 DR Pfam; PF01376; HCV NS2; 1.
 DR Pfam; PF01375; HCV NS3; 1.
 DR Pfam; PF01374; HCV NS4a; 1.
 DR Pfam; PF01373; HCV NS4b; 1.
 DR Pfam; PF01372; HCV NS5a; 1.
 DR Pfam; PF01371; HCV RdRp; 1.
 DR Pfam; PF01370; HCV NS1; 1.
 DR Pfam; PF01369; HCV NS2; 1.
 DR Pfam; PF01368; HCV NS3; 1.
 DR Pfam; PF01367; HCV NS4a; 1.
 DR Pfam; PF01366; HCV NS4b; 1.
 DR Pfam; PF01365; HCV NS5a; 1.
 DR Pfam; PF01364; HCV RdRp; 1.
 DR Pfam; PF01363; HCV NS1; 1.
 DR Pfam; PF01362; HCV NS2; 1.
 DR Pfam; PF01361; HCV NS3; 1.
 DR Pfam; PF01360; HCV NS4a; 1.
 DR Pfam; PF01359; HCV NS4b; 1.
 DR Pfam; PF01358; HCV NS5a; 1.
 DR Pfam; PF01357; HCV RdRp; 1.
 DR Pfam; PF01356; HCV NS1; 1.
 DR Pfam; PF01355; HCV NS2; 1.
 DR Pfam; PF01354; HCV NS3; 1.
 DR Pfam; PF01353; HCV NS4a; 1.
 DR Pfam; PF01352; HCV NS4b; 1.
 DR Pfam; PF01351; HCV NS5a; 1.
 DR Pfam; PF01350; HCV RdRp; 1.
 DR Pfam; PF01349; HCV NS1; 1.
 DR Pfam; PF01348; HCV NS2; 1.
 DR Pfam; PF01347; HCV NS3; 1.
 DR Pfam; PF01346; HCV NS4a; 1.
 DR Pfam; PF01345; HCV NS4b; 1.
 DR Pfam; PF01344; HCV NS5a; 1.
 DR Pfam; PF01343; HCV RdRp; 1.
 DR Pfam; PF01342; HCV NS1; 1.
 DR Pfam; PF01341; HCV NS2; 1.
 DR Pfam; PF01340; HCV NS3; 1.
 DR Pfam; PF01339; HCV NS4a; 1.
 DR Pfam; PF01338; HCV NS4b; 1.
 DR Pfam; PF01337; HCV NS5a; 1.
 DR Pfam; PF01336; HCV RdRp; 1.
 DR Pfam; PF01335; HCV NS1; 1.
 DR Pfam; PF01334; HCV NS2; 1.
 DR Pfam; PF01333; HCV NS3; 1.
 DR Pfam; PF01332; HCV NS4a; 1.
 DR Pfam; PF01331; HCV NS4b; 1.
 DR Pfam; PF01330; HCV NS5a; 1.
 DR Pfam; PF01329; HCV RdRp; 1.
 DR Pfam; PF01328; HCV NS1; 1.
 DR Pfam; PF01327; HCV NS2; 1.
 DR Pfam; PF01326; HCV NS3; 1.
 DR Pfam; PF01325; HCV NS4a; 1.
 DR Pfam; PF01324; HCV NS4b; 1.
 DR Pfam; PF01323; HCV NS5a; 1.
 DR Pfam; PF01322; HCV RdRp; 1.
 DR Pfam; PF01321; HCV NS1; 1.
 DR Pfam; PF01320; HCV NS2; 1.
 DR Pfam; PF01319; HCV NS3; 1.
 DR Pfam; PF01318; HCV NS4a; 1.
 DR Pfam; PF01317; HCV NS4b; 1.
 DR Pfam; PF01316; HCV NS5a; 1.
 DR Pfam; PF01315; HCV RdRp; 1.
 DR Pfam; PF01314; HCV NS1; 1.
 DR Pfam; PF01313; HCV NS2; 1.
 DR Pfam; PF01312; HCV NS3; 1.
 DR Pfam; PF01311; HCV NS4a; 1.
 DR Pfam; PF01310; HCV NS4b; 1.
 DR Pfam; PF01309; HCV NS5a; 1.
 DR Pfam; PF01308; HCV RdRp; 1.
 DR Pfam; PF01307; HCV NS1; 1.
 DR Pfam; PF01306; HCV NS2; 1.
 DR Pfam; PF01305; HCV NS3; 1.
 DR Pfam; PF01304; HCV NS4a; 1.
 DR Pfam; PF01303; HCV NS4b; 1.
 DR Pfam; PF01302; HCV NS5a; 1.
 DR Pfam; PF01301; HCV RdRp; 1.
 DR Pfam; PF01300; HCV NS1; 1.
 DR Pfam; PF01299; HCV NS2; 1.
 DR Pfam; PF01298; HCV NS3; 1.
 DR Pfam; PF01297; HCV NS4a; 1.
 DR Pfam; PF01296; HCV NS4b; 1.
 DR Pfam; PF01295; HCV NS5a; 1.
 DR Pfam; PF01294; HCV RdRp; 1.
 DR Pfam; PF01293; HCV NS1; 1.
 DR Pfam; PF01292; HCV NS2; 1.
 DR Pfam; PF01291; HCV NS3; 1.
 DR Pfam; PF01290; HCV NS4a; 1.
 DR Pfam; PF01289; HCV NS4b; 1.
 DR Pfam; PF01288; HCV NS5a; 1.
 DR Pfam; PF01287; HCV RdRp; 1.
 DR Pfam; PF01286; HCV NS1; 1.
 DR Pfam; PF01285; HCV NS2; 1.
 DR Pfam; PF01284; HCV NS3; 1.
 DR Pfam; PF01283; HCV NS4a; 1.
 DR Pfam; PF01282; HCV NS4b; 1.
 DR Pfam; PF01281; HCV NS5a; 1.
 DR Pfam; PF01280; HCV RdRp; 1.
 DR Pfam; PF01279; HCV NS1; 1.
 DR Pfam; PF01278; HCV NS2; 1.
 DR Pfam; PF01277; HCV NS3; 1.
 DR Pfam; PF01276; HCV NS4a; 1.
 DR Pfam; PF01275; HCV NS4b; 1.
 DR Pfam; PF01274; HCV NS5a; 1.
 DR Pfam; PF01273; HCV RdRp; 1.
 DR Pfam; PF01272; HCV NS1; 1.
 DR Pfam; PF01271; HCV NS2; 1.
 DR Pfam; PF01270; HCV NS3; 1.
 DR Pfam; PF01269; HCV NS4a; 1.
 DR Pfam; PF01268; HCV NS4b; 1.
 DR Pfam; PF01267; HCV NS5a; 1.
 DR Pfam; PF01266; HCV RdRp; 1.
 DR Pfam; PF01265; HCV NS1; 1.
 DR Pfam; PF01264; HCV NS2; 1.
 DR Pfam; PF01263; HCV NS3; 1.
 DR Pfam; PF01262; HCV NS4a; 1.
 DR Pfam; PF01261; HCV NS4b; 1.
 DR Pfam; PF01260; HCV NS5a; 1.
 DR Pfam; PF01259; HCV RdRp; 1.
 DR Pfam; PF01258; HCV NS1; 1.
 DR Pfam; PF01257; HCV NS2; 1.
 DR Pfam; PF01256; HCV NS3; 1.
 DR Pfam; PF01255; HCV NS4a; 1.
 DR Pfam; PF01254; HCV NS4b; 1.
 DR Pfam; PF01253; HCV NS5a; 1.
 DR Pfam; PF01252; HCV RdRp; 1.
 DR Pfam; PF01251; HCV NS1; 1.
 DR Pfam; PF01250; HCV NS2; 1.
 DR Pfam; PF01249; HCV NS3; 1.
 DR Pfam; PF01248; HCV NS4a; 1.
 DR Pfam; PF01247; HCV NS4b; 1.
 DR Pfam; PF01246; HCV NS5a; 1.
 DR Pfam; PF01245; HCV RdRp; 1.
 DR Pfam; PF01244; HCV NS1; 1.
 DR Pfam; PF01243; HCV NS2; 1.
 DR Pfam; PF01242; HCV NS3; 1.
 DR Pfam; PF01241; HCV NS4a; 1.
 DR Pfam; PF01240; HCV NS4b; 1.
 DR Pfam; PF01239; HCV NS5a; 1.
 DR Pfam; PF01238; HCV RdRp; 1.
 DR Pfam; PF01237; HCV NS1; 1.
 DR Pfam; PF01236; HCV NS2; 1.
 DR Pfam; PF01235; HCV NS3; 1.
 DR Pfam; PF01234; HCV NS4a; 1.
 DR Pfam; PF01233; HCV NS4b; 1.
 DR Pfam; PF01232; HCV NS5a; 1.
 DR Pfam; PF01231; HCV RdRp; 1.
 DR Pfam; PF01230; HCV NS1; 1.
 DR Pfam; PF01229; HCV NS2; 1.
 DR Pfam; PF01228; HCV NS3; 1.
 DR Pfam; PF01227; HCV NS4a; 1.
 DR Pfam; PF01226; HCV NS4b; 1.
 DR Pfam; PF01225; HCV NS5a; 1.
 DR Pfam; PF01224; HCV RdRp; 1.
 DR Pfam; PF01223; HCV NS1; 1.
 DR Pfam; PF01222; HCV NS2; 1.
 DR Pfam; PF01221; HCV NS3; 1.
 DR Pfam; PF01220; HCV NS4a; 1.
 DR Pfam; PF01219; HCV NS4b; 1.
 DR Pfam; PF01218; HCV NS5a; 1.
 DR Pfam; PF01217; HCV RdRp; 1.
 DR Pfam; PF01216; HCV NS1; 1.
 DR Pfam; PF01215; HCV NS2; 1.
 DR Pfam; PF01214; HCV NS3; 1.
 DR Pfam; PF01213; HCV NS4a; 1.
 DR Pfam; PF01212; HCV NS4b; 1.
 DR Pfam; PF01211; HCV NS5a; 1.
 DR Pfam; PF01210; HCV RdRp; 1.
 DR Pfam; PF01209; HCV NS1; 1.
 DR Pfam; PF01208; HCV NS2; 1.
 DR Pfam; PF01207; HCV NS3; 1.
 DR Pfam; PF01206; HCV NS4a; 1.
 DR Pfam; PF01205; HCV NS4b; 1.
 DR Pfam; PF01204; HCV NS5a; 1.
 DR Pfam; PF01203; HCV RdRp; 1.
 DR Pfam; PF01202; HCV NS1; 1.
 DR Pfam; PF01201; HCV NS2; 1.
 DR Pfam; PF01200; HCV NS3; 1.
 DR Pfam; PF01199; HCV NS4a; 1.
 DR Pfam; PF01198; HCV NS4b; 1.
 DR Pfam; PF01197; HCV NS5a; 1.
 DR Pfam; PF01196; HCV RdRp; 1.
 DR Pfam; PF01195; HCV NS1; 1.
 DR Pfam; PF01194; HCV NS2; 1.
 DR Pfam; PF01193; HCV NS3; 1.
 DR Pfam; PF01192; HCV NS4a; 1.
 DR Pfam; PF01191; HCV NS4b; 1.
 DR Pfam; PF01190; HCV NS5a; 1.
 DR Pfam; PF01189; HCV RdRp; 1.
 DR Pfam; PF01188; HCV NS1; 1.
 DR Pfam; PF01187; HCV NS2; 1.
 DR Pfam; PF01186; HCV NS3; 1.
 DR Pfam; PF01185; HCV NS4a; 1.
 DR Pfam; PF01184; HCV NS4b; 1.
 DR Pfam; PF01183; HCV NS5a; 1.
 DR Pfam; PF01182; HCV RdRp; 1.
 DR Pfam; PF01181; HCV NS1; 1.
 DR Pfam; PF01180; HCV NS2; 1.
 DR Pfam; PF01179; HCV NS3; 1.
 DR Pfam; PF01178; HCV NS4a; 1.
 DR Pfam; PF01177; HCV NS4b; 1.
 DR Pfam; PF01176; HCV NS5a; 1.
 DR Pfam; PF01175; HCV RdRp; 1.
 DR Pfam; PF01174; HCV NS1; 1.
 DR Pfam; PF01173; HCV NS2; 1.
 DR Pfam; PF01172; HCV NS3; 1.
 DR Pfam; PF01171; HCV NS4a; 1.
 DR Pfam; PF01170; HCV NS4b; 1.
 DR Pfam; PF01169; HCV NS5a; 1.
 DR Pfam; PF01168; HCV RdRp; 1.
 DR Pfam; PF01167; HCV NS1; 1.
 DR Pfam; PF01166; HCV NS2; 1.
 DR Pfam; PF01165; HCV NS3; 1.
 DR Pfam; PF01164; HCV NS4a; 1.
 DR Pfam; PF01163; HCV NS4b; 1.
 DR Pfam; PF01162; HCV NS5a; 1.
 DR Pfam; PF01161; HCV RdRp; 1.
 DR Pfam; PF01160; HCV NS1; 1.
 DR Pfam; PF01159; HCV NS2; 1.
 DR Pfam; PF01158; HCV NS3; 1.
 DR Pfam; PF01157; HCV NS4a; 1.
 DR Pfam; PF01156; HCV NS4b; 1.
 DR Pfam; PF01155; HCV NS5a; 1.
 DR Pfam; PF01154; HCV RdRp; 1.
 DR Pfam; PF01153; HCV NS1; 1.
 DR Pfam; PF01152; HCV NS2; 1.
 DR Pfam; PF01151; HCV NS3; 1.
 DR Pfam; PF01150; HCV NS4a; 1.
 DR Pfam; PF01149; HCV NS4b; 1.
 DR Pfam; PF01148; HCV NS5a; 1.
 DR Pfam; PF01147; HCV RdRp; 1.
 DR Pfam; PF01146; HCV NS1; 1.
 DR Pfam; PF01145; HCV NS2; 1.
 DR Pfam; PF01144; HCV NS3; 1.
 DR Pfam; PF01143; HCV NS4a; 1.
 DR Pfam; PF01142; HCV NS4b; 1.
 DR Pfam; PF01141; HCV NS5a; 1.
 DR Pfam; PF01140; HCV RdRp; 1.
 DR Pfam; PF01139; HCV NS1; 1.
 DR Pfam; PF01138; HCV NS2; 1.
 DR Pfam; PF01137; HCV NS3; 1.
 DR Pfam; PF01136; HCV NS4a; 1.
 DR Pfam; PF01135; HCV NS4b; 1.
 DR Pfam; PF01134; HCV NS5a; 1.
 DR Pfam; PF01133; HCV RdRp; 1.
 DR Pfam; PF01132; HCV NS1; 1.
 DR Pfam; PF01131; HCV NS2; 1.
 DR Pfam; PF01130; HCV NS3; 1.
 DR Pfam; PF01129; HCV NS4a; 1.
 DR Pfam; PF01128; HCV NS4b; 1.
 DR Pfam; PF01127; HCV NS5a; 1.
 DR Pfam; PF01126; HCV RdRp; 1.
 DR Pfam; PF01125; HCV NS1; 1.
 DR Pfam; PF01124; HCV NS2; 1.
 DR Pfam; PF01123; HCV NS3; 1.
 DR Pfam; PF01122; HCV NS4a; 1.
 DR Pfam; PF01121; HCV NS4b; 1.
 DR Pfam; PF01120; HCV NS5a; 1.
 DR Pfam; PF01119; HCV RdRp; 1.
 DR Pfam; PF01118; HCV NS1; 1.
 DR Pfam; PF01117; HCV NS2; 1.
 DR Pfam; PF01116; HCV NS3; 1.
 DR Pfam; PF01115; HCV NS4a; 1.
 DR Pfam; PF01114; HCV NS4b; 1.
 DR Pfam; PF01113; HCV NS5a; 1.
 DR Pfam; PF01112; HCV RdRp; 1.
 DR Pfam; PF01111; HCV NS1; 1.
 DR Pfam; PF01110; HCV NS2; 1.
 DR Pfam; PF01109; HCV NS3; 1.
 DR Pfam; PF01108; HCV NS4a; 1.
 DR Pfam; PF01107; HCV NS4b; 1.
 DR Pfam; PF01106; HCV NS5a; 1.
 DR Pfam; PF01105; HCV RdRp; 1.
 DR Pfam; PF01104; HCV NS1; 1.
 DR Pfam; PF01103; HCV NS2; 1.
 DR Pfam; PF01102; HCV NS3; 1.
 DR Pfam; PF01101; HCV NS4a; 1.
 DR Pfam; PF01100; HCV NS4b; 1.
 DR Pfam; PF01099; HCV NS5a; 1.
 DR Pfam; PF01098; HCV RdRp; 1.
 DR Pfam; PF01097; HCV NS1; 1.
 DR Pfam; PF01096; HCV NS2; 1.
 DR Pfam; PF01095; HCV NS3; 1.
 DR Pfam; PF01094; HCV NS4a; 1.
 DR Pfam; PF01093; HCV NS4b; 1.
 DR Pfam; PF01092; HCV NS5a; 1.
 DR Pfam; PF01091; HCV RdRp; 1.
 DR Pfam; PF01090; HCV NS1; 1.
 DR Pfam; PF01089; HCV NS2; 1.
 DR Pfam; PF01088; HCV NS3; 1.
 DR Pfam; PF01087; HCV NS4a; 1.
 DR Pfam; PF01086; HCV NS4b; 1.
 DR Pfam; PF01085; HCV NS5a; 1.
 DR Pfam; PF01084; HCV RdRp; 1.
 DR Pfam; PF01083; HCV NS1; 1.
 DR Pfam; PF01082; HCV NS2; 1.
 DR Pfam; PF01081; HCV NS3; 1.
 DR Pfam; PF01080; HCV NS4a; 1.
 DR Pfam; PF01079; HCV NS4b; 1.
 DR Pfam; PF01078; HCV NS5a; 1.
 DR Pfam; PF01077; HCV RdRp; 1.
 DR Pfam; PF01076; HCV NS1; 1.
 DR Pfam; PF01075; HCV NS2; 1.
 DR Pfam; PF01074; HCV NS3; 1.
 DR Pfam; PF01073; HCV NS4a; 1.
 DR Pfam; PF01072; HCV NS4b; 1.
 DR Pfam; PF01071; HCV NS5a; 1.
 DR Pfam; PF01070; HCV RdRp; 1.
 DR Pfam; PF01069; HCV NS1; 1.
 DR Pfam; PF01068; HCV NS2; 1.
 DR Pfam; PF01067; HCV NS3; 1.
 DR Pfam; PF01066; HCV NS4a; 1.
 DR Pfam; PF01065; HCV NS4b; 1.
 DR Pfam; PF01064; HCV NS5a; 1.
 DR Pfam; PF01063; HCV RdRp; 1.
 DR Pfam; PF01062; HCV NS1; 1.
 DR Pfam; PF01061; HCV NS2; 1.
 DR Pfam; PF01060; HCV NS3; 1.
 DR Pfam; PF01059; HCV NS4a; 1.
 DR Pfam; PF01058; HCV NS4b; 1.
 DR Pfam; PF01057; HCV NS5a; 1.
 DR Pfam; PF01056; HCV RdRp; 1.

QY 5 NYTIFKIRMYVGVEH 20
|||||
Db 627 NYTIFKIRMYVGVEH 642

RESULT 11

Q9QF35 PRELIMINARY; PRT; 3033 AA.
AC Q9QF35;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J6CH;
RX MEDLINE=99420396; PubMed=10489358;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Hepatitis C virus: an infectious molecular clone of a second major
RT genotype (2a) and lack of viability of intertypic 1a and 2a
RT chimeras.";
RL Virology 262:250-263(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J6CH;
RA Bukh J.;
RT Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF17036; AAF01178.1; -;
DR HSP; P27958; IHEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4b.
DR InterPro; IPR001490; HCV NS4a.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR002129; Pyridoxal deC.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01538; HCV NS1; 1.
DR Pfam; PF01537; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00392; DDC GAD HDC YDC; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05052; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; transferase; Transmembrane.
SQ SEQUENCE 3033 AA; 329263 MW; 6D894082FF85C7 CRC64;

Query Match 80.0%; Score 16; DB 12; Length 3033;
Best Local Similarity 100.0%; Pred. No. 4e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 NYTIFKIRMYVGVEH 20
|||||
Db 627 NYTIFKIRMYVGVEH 642

RESULT 12

Q91ZA3 PRELIMINARY; PRT; 3033 AA.
AC Q91ZA3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD2a-5;
RA Itakura J., Nagayama K., Enomoto N., Kurosaki M., Watanabe H.,
RA Sato C.;
RT "Full length cDNA sequence of HCV genotype 2a, strain MD2a-5.";
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF238484; AAF59943.1; -;
DR HSP; P27958; IHEI.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002519; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01538; HCV NS1; 1.
DR Pfam; PF01537; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD186062; HCV NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05052; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; transferase; Transmembrane.
SQ SEQUENCE 3033 AA; 329515 MW; E016D0E175644593 CRC64;

Query Match 80.0%; Score 16; DB 12; Length 3033;
Best Local Similarity 100.0%; Pred. No. 4e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 NYTIFKIRMYVGVEH 20
|||||
Db 627 NYTIFKIRMYVGVEH 642

Query Match 80.0%; Score 16; DB 12; Length 424;
 Best Local Similarity 100.0%; Pred. No. 7e-10; Indels 0; Gaps 0;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 NYTIFKIRMYGGVEH 20
 DB 286 NYTIFKIRMYGGVEH 301

RESULT 7

Q9PWX5 PRELIMINARY; PRT; 3015 AA.
 AC Q9PWX5
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99420396; PubMed=10489358;
 RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
 RT "Hepatitis C virus: an infectious molecular clone of a second major
 RT genotype (2a) and lack of viability of intertypic 1a and 2a
 RT chimeras";
 RL Virology 262:250-263 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Bukh J.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL; AF177040; AAF01182.1; -;
 DR EMBL; AF177038; AAF01180.1; -;
 DR HSP; P27958; 1HEI.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR002129; Pyridoxal dec.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase C; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS00392; DDC GAD HDC YDC; 1.
 DR PROSITE; PS05057; RDRP POSITIVE; 1.
 DR PROSITE; PS05021; RDRP VIRAL; 1.
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydrolase; Nonstructural protein; Polyprotein;

KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3015 AA; 328159 MW; B7D23BC1F190663A CRC64;

Query Match 80.0%; Score 16; DB 12; Length 3015;
 Best Local Similarity 100.0%; Pred. No. 4e-09;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 NYTIFKIRMYGGVEH 20
 DB 627 NYTIFKIRMYGGVEH 642

RESULT 8

Q9PWU9 PRELIMINARY; PRT; 3015 AA.
 AC Q9PWU9
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99420396; PubMed=10489358;
 RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
 RT "Hepatitis C virus: an infectious molecular clone of a second major
 RT genotype (2a) and lack of viability of intertypic 1a and 2a
 RT chimeras";
 RL Virology 262:250-263 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Bukh J.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL; AF177039; AAF01181.1; -;
 DR EMBL; AF177037; AAF01179.1; -;
 DR HSP; P27958; 1HEI.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR002129; Pyridoxal dec.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase C; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS00392; DDC GAD HDC YDC; 1.
 DR PROSITE; PS05057; RDRP POSITIVE; 1.
 DR PROSITE; PS05021; RDRP VIRAL; 1.
 DR PROSITE; PS05057; RDRP POSITIVE; 1.

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NDM59;
RX MEDLINE=21361470; PubMed=11468731;
RA Kurihara C., Ishiyama N., Nishiyama Y., Fukushi S., Kageyama T.,
RA Katayama K., Miura S.;
RT "Molecular characterization of hepatitis C virus genotype 2a from the
RT entire sequences of four isolates."
RL J. Med. Virol. 64:466-475(2001).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF169005; AAF25613.1; -;
DR HSSP; P27958; 1HEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC1; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05051; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3033 AA; 329226 MW; 21492388CA0D5D8C CRC64;

Query Match 100.0%; Score 20; DB 12; Length 3033;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINVTIFKIRMYGVGVH 20
Db 623 PCTINVTIFKIRMYGVGVH 642

RESULT 5
O37937
ID O37937 PRELIMINARY; PRT; 364 AA.
AC O37937;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE E2/NS1 protein (Genome polyprotein) (Fragment).
GN E2/NS1
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-W14;
RX MEDLINE=21028803; PubMed=11155671;
RA Wu C., Tao Q.;
RT "Comparison between homologies of E2/NS1 gene from genotype III
RT Chinese isolates of hepatitis C virus and that from reported
RT isolates."
RL Chin. Med. J. 111:807-809(1998).
DR EMBL; AF023154; AAC61802.1; -;
DR InterPro; IPR002531; HCV_NS1.
DR Pfam; PF01560; HCV_NS1; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 364 364
SQ SEQUENCE 364 AA; 40610 MW; 4C160290877A9A0 CRC64;

Query Match 80.0%; Score 16; DB 12; Length 364;
Best Local Similarity 100.0%; Pred. No. 6.1e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 NYTIFKIRMYGVGVH 20
Db 250 NYTIFKIRMYGVGVH 265

RESULT 6
Q98UN6
ID Q98UN6 PRELIMINARY; PRT; 424 AA.
AC Q98UN6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Q2a;
RX MEDLINE=20499063; PubMed=11044085;
RA Hadlock K.G., Lanford R.E., Perkins S., Rowe J., Yang Q., Levy S.,
RA Pileri P., Abrighani S., Fong S.K.;
RT "Human monoclonal antibodies that inhibit binding of hepatitis C virus
RT E2 protein to CD81 and recognize conserved conformational epitopes."
RL J. Virol. 74:10407-10416(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Q2a;
RA Keck Z.Y., Hadlock K.G., Yang Q.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF348702; AAK32683.1; -;
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT CHAIN 6 >42 E1.
FT CHAIN 43 >408 E2.
FT CHAIN 409 >424 P7.
FT NON_TER 424 424
SQ SEQUENCE 424 AA; 46996 MW; 458E648991A7749A CRC64;

```
FT CHAIN 401 >420 P7.
FT NON_TER 420 420
SQ SEQUENCE 420 AA; 45882 MW; F28D0C3EE0CD081E CRC64;

Query Match 100.0%; Score 20; DB 12; Length 420;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PCTINYTIKIRMYVGGVEH 20
Db 273 PCTINYTIKIRMYVGGVEH 292

RESULT 2
ID Q81756 PRELIMINARY; PRT; 2436 AA.
AC Q81756;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
ON NCBI_TaxID=11103;
RX MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
sequence.";
RL J. Gen. Virol. 82:1291-1297(2001).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
DR EMBL; AF271632; AAF81759.1; -.
DR HSSP; P27958; 1A1V.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_Core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002866; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXdc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05052; RDRP_VIRAL; 1.
DR ATP-binding; Coa protein; Envelope protein; Glycoprotein; Helicase;
KW Hydroxylase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327124 MW; 2489CE74AC864E58 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PCTINYTIKIRMYVGGVEH 20
Db 619 PCTINYTIKIRMYVGGVEH 638

RESULT 4
Q909A7
ID Q909A7 PRELIMINARY; PRT; 3033 AA.
AC Q909A7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:45:57 ; Search time 24.4 Seconds
(without alignments)
211.519 Million cell updates/sec

Title: US-09-973-025-89

Perfect score: 20

Sequence: 1 PQTINYTIKIRMYGVGEH 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- SPREMBL_23.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phase.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	420	Q98UN4	Q98un4 hepatitis c
2	20	100.0	2436	Q81756	Q81756 hepatitis c
3	20	100.0	3011	Q91F55	Q91f55 hepatitis c
4	20	100.0	3033	Q909A7	Q909a7 hepatitis c
5	16	80.0	364	O37937	O37937 hepatitis c
6	16	80.0	424	Q98UN6	Q98un6 hepatitis c
7	16	80.0	3015	Q9PMX5	Q9pmx5 hepatitis c
8	16	80.0	3015	Q9PMU9	Q9pmu9 hepatitis c
9	16	80.0	3033	Q9Q9A8	Q9q9a8 hepatitis c
10	16	80.0	3033	Q99IB6	Q99ib6 hepatitis c
11	16	80.0	3033	Q9QF35	Q9qf35 hepatitis c
12	16	80.0	3033	Q91Z33	Q91z33 hepatitis c
13	16	80.0	3033	Q91Z34	Q91z34 hepatitis c
14	16	80.0	3033	Q9Q9B0	Q9q9b0 hepatitis c
15	14	70.0	133	Q81497	Q81497 hepatitis c
16	14	70.0	133	Q81506	Q81506 hepatitis c

17	14	70.0	133	12	Q81539	Q81539 hepatitis c
18	14	70.0	133	12	Q81509	Q81509 hepatitis c
19	14	70.0	133	12	Q81515	Q81515 hepatitis c
20	14	70.0	133	12	Q81527	Q81527 hepatitis c
21	14	70.0	133	12	Q81530	Q81530 hepatitis c
22	14	70.0	133	12	Q81533	Q81533 hepatitis c
23	14	70.0	133	12	Q81536	Q81536 hepatitis c
24	14	70.0	133	12	Q81500	Q81500 hepatitis c
25	14	70.0	133	12	Q81524	Q81524 hepatitis c
26	14	70.0	133	12	Q81503	Q81503 hepatitis c
27	14	70.0	133	12	Q81518	Q81518 hepatitis c
28	14	70.0	782	12	Q68953	Q68953 hepatitis c
29	14	70.0	783	12	Q68952	Q68952 hepatitis c
30	14	70.0	3010	12	Q68949	Q68949 hepatitis c
31	14	70.0	3010	12	Q9J3G0	Q9j3g0 hepatitis c
32	14	70.0	3032	12	Q99IB4	Q99ib4 hepatitis c
33	14	70.0	3033	12	Q99IB8	Q99ib8 hepatitis c
34	13	65.0	133	12	Q81521	Q81521 hepatitis c
35	13	65.0	783	12	Q68948	Q68948 hepatitis c
36	13	65.0	3032	12	Q99IB2	Q99ib2 hepatitis c
37	13	65.0	3032	12	Q9Q9A9	Q9q9a9 hepatitis c
38	13	65.0	3033	12	Q91Z35	Q91z35 hepatitis c
39	12	60.0	129	12	Q91ED2	Q91ed2 hepatitis c
40	12	60.0	129	12	Q91ED5	Q91ed5 hepatitis c
41	12	60.0	129	12	Q91EG1	Q91eg1 hepatitis c
42	12	60.0	129	12	Q91EG0	Q91eg0 hepatitis c
43	12	60.0	746	12	Q8JPM2	Q8jpm2 hepatitis c
44	12	60.0	782	12	Q68951	Q68951 hepatitis c
45	12	60.0	3013	12	Q9QIX9	Q9qix9 hepatitis c

ALIGNMENTS

RESULT 1

Q98UN4 ID Q98UN4 PRELIMINARY; PRT; 420 AA.
AC Q98UN4;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Q1a;
RX MEDLINE=20499063; PubMed=11044085;
RA Hadlock K.G., Lanford R.E., Perkins S., Rowe J., Yang Q., Levy S.,
Pileri P., Abrignani S., Fong S.K.;
RT "Human monoclonal antibodies that inhibit binding of hepatitis C virus
E2 protein to CD81 and recognize conserved conformational epitopes.";
J. Virol. 74:10407-10416(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Q1a;
RA Keck Z.Y., Hadlock K.G., Yang Q.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF348704; AAK32685.1; -
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV_NSI.
DR InterPro; IPR002518; HCV_NS2.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR -ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
Polyprotein; Transmembrane.
FT NON_TER 1
FT CHAIN 1 >37 E1.
FT CHAIN 38 >400 E2.

THIS PAGE BLANK (USPTO)

Query Match 30.0%; Score 6; DB 1; Length 265;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 VGGVEH 20
| | | | |
Db 31 VGGVEH 36

Search completed: November 21, 2003, 21:00:02
Job time : 4.85 secs

RT "Phosphorylation of maize RAB-17 protein by casein kinase 2.";
 RL J. Biol. Chem. 266:22510-22514(1991).
 CC -1- DEVELOPMENTAL STAGE: ABA TREATMENT INDUCED THE SYNTHESIS OF RAB-17
 CC IN CALLI, HOWEVER, THE RAB-17 PROTEINS WERE FOUND TO BE HIGHLY
 CC PHOSPHORYLATED ONLY IN EMBRYOS.
 CC -1- INDUCTION: By abscisic acid (ABA) and water stress.
 CC -1- PTM: SERINE IS THE ONLY AMINO ACID WHICH BECOMES PHOSPHORYLATED.
 CC -1- SIMILARITY: BELONGS TO THE PLANT DEHYDRIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X15290; CAA33364.1; -;
 CC EMBL; X15994; CAA34123.1; -;
 CC PIR; A39316; A39316.
 CC PIR; S05545; S05545.
 CC PIR; S08633; S08633.
 CC MaizeDB; 24916; -;
 CC InterPro; IPR000167; Dehydrin.
 CC Pfam; PF00257; dehydrin; 1.
 CC PROSITE; PS00315; DEHYDRIN_1; 1.
 CC PROSITE; PS00823; DEHYDRIN_2; 2.
 CC Dehydrin; Repeat; Phosphorylation; Multigene family.
 CC FT DOMAIN 76 84 POLY-SER.
 CC FT REPEAT 2 10 1-1.
 CC FT REPEAT 122 130 1-2.
 CC FT REPEAT 132 140 1-3.
 CC FT DOMAIN 94 166 2 X APPROXIMATE REPEATS.
 CC FT REPEAT 94 107 2-1.
 CC FT REPEAT 153 166 2-2.
 CC FT CONFLICT 9 9 R -> H (IN REF. 2).
 CC FT CONFLICT 13 13 R -> A (IN REF. 2).
 CC FT CONFLICT 41 41 MISSING (IN REF. 2).
 CC SQ SEQUENCE 168 AA; 17161 MW; BE7758CAD37CFF39 CRC64;
 CC
 CC Query Match 30.0%; Score 6; DB 1; Length 168;
 CC Best Local Similarity 100.0%; Pred. No. 8.3;
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 15 VGVGEH 20
 CC Db 24 VGVGEH 29
 CC
 CC RESULT 14
 CC YGZA YEAST STANDARD; PRT; 245 AA.
 CC AC PS3060;
 CC DT 01-OCT-1996 (Rel. 34, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 01-NOV-1997 (Rel. 35, Last annotation update)
 CC DE Hypothetical 28.2 kDa protein in HFW1-PDE1 intergenic region.
 CC GN YGL250W OR NRC245.
 CC OS Saccharomyces cerevisiae (Baker's yeast).
 CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 CC OX NCBI_TaxID=4932;
 CC RN [1]
 CC RN SEQUENCE FROM N.A.
 CC RP STRAIN=S288C / FY1679;
 CC RX MEDLINE=97127827; PubMed=89725578;
 CC RA Coissac E., Maillier E., Robineau S., Netter P.;
 CC RT "Sequence of a 39,411 bp DNA fragment covering the left end of
 CC chromosome VII of Saccharomyces cerevisiae.";
 CC RL Yeast 12:1555-1562(1996).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X94357; CAA64137.1; -;
 CC EMBL; Z72772; CAA96970.1; -;
 CC PIR; S61611; S61611.
 CC SGD; S0003219; YGL250W.
 CC KW Hypothetical protein.
 CC SQ SEQUENCE 245 AA; 28184 MW; 0C936065C5C6524B CRC64;
 CC
 CC Query Match 30.0%; Score 6; DB 1; Length 245;
 CC Best Local Similarity 100.0%; Pred. No. 11;
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 6 YTIFFI 11
 CC Db 151 YTIFFI 156
 CC
 CC RESULT 15
 CC DABP AQUAE STANDARD; PRT; 265 AA.
 CC AC O67061;
 CC DT 15-JUL-1999 (Rel. 38, Created)
 CC DT 15-JUL-1999 (Rel. 38, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Dihydropicolinate reductase (EC 1.3.1.26) (DHPR).
 CC GN DABP OR AQ_916.
 CC OS Aquifex aeolicus.
 CC OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
 CC OX NCBI_TaxID=63363;
 CC RN [1]
 CC RN SEQUENCE FROM N.A.
 CC RC STRAIN=VF5;
 CC RX MEDLINE=98196666; PubMed=9537320;
 CC RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L., R.,
 CC Graham D.E., Overbeek R., Snead M.A., Aujaay M., Huber R.,
 CC Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 CC "The complete genome of the hyperthermophilic bacterium Aquifex
 CC aeolicus".
 CC RL Nature 392:353-358(1998).
 CC CC -1- CATALYTIC ACTIVITY: 2,3,4,5-tetrahydropicolinate + NAD(P)(+) =
 CC 2,3-dihydropicolinate + NAD(P)H.
 CC CC -1- PATHWAY: Biosynthesis of diaminopimelate and lysine from aspartate
 CC semialdehyde; second step.
 CC CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC CC -1- SIMILARITY: Belongs to the dihydropicolinate reductase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE000713; AAC07008.1; -;
 CC PIR; B70379; B70379.
 CC HSSP; P04036; IDRW.
 CC HAMAP; MF_00102; -; 1.
 CC InterPro; IPR000846; DapB.
 CC Pfam; PF05173; DapB_C; 1.
 CC Pfam; PF01113; DapB_N; 1.
 CC ProDom; PD004105; DapB; 1.
 CC TIGRfam; TIGR00036; dapB; 1.
 CC PROSITE; PS01298; DAPB; 1.
 CC KW Diaminopimelate biosynthesis; Lysine biosynthesis; Oxidoreductase;
 CC NADP; Complete proteome.
 CC SQ SEQUENCE 265 AA; 28960 MW; EC3240CBCE8BE27A CRC64;


```
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 45.0%; Score 9; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 0.052;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 RMYVGVEH 20
DB 630 RMYVGVEH 638

RESULT 10
POLG_HCVJ8 STANDARD; PRT; 3033 AA.
AC P26661;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J8) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11115;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
RA Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC -!- HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA) (N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY_S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D10988; BAA01761.1; -
CC PIR; A40250; GNMVJ8.
CC HSSP; P27956; IHEI.
CC MEROPS; S29.001; -.
CC MEROPS; U39.001; -.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002522; HCV capsid.
CC InterPro; IPR002521; HCV_core.
```

```
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV_NSI.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR SMART; SM00487; DEXDc; 1.
DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 733 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 734 1010 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT CHAIN 1011 1619 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1620 1866 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1867 2017 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 2018 3033 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT ACT_SITE 1087 1087 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1111 1111 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1169 1169 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1234 1241 ATP (POTENTIAL).
FT SITE 1320 1323 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2359 2359 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3981FD1A CRC64;

Query Match 45.0%; Score 9; DB 1; Length 3033;
Best Local Similarity 100.0%; Pred. No. 0.052;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 RMYVGVEH 20
DB 634 RMYVGVEH 642
```

FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;

Query Match
 Best Local Similarity 45.0%; Score 9; DB 1; Length 3010;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 RMYGVGVH 20
 |||||
 Db 630 RMYGVGVH 638

RESULT 9
 POLG_HCVTW STANDARD; PRT; 3010 AA.
 AC P29846:
 DT 01-APR-1993 (Rel. 25, Created)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230206; PubMed=1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 RT "The Taiwanese hepatitis C virus genome: sequence determination and
 mapping the 5' terminus of viral genome and antigenomic RNA.";
 RL Virology 188:102-113(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA}(N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).

 EMBL; M84754; -, NOT_ANNOTATED_CDS.
 PIR; A40244; GNWVTV.
 PDB; 1N64; 25-FEB-03.
 PDB; 1NS3; 08-APR-98.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure. 1
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CORE PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 3010 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).


```

FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1031 1035 HELIX
FT STRAND 1039 1047 STRAND
FT STRAND 1050 1050 STRAND
FT STRAND 1059 1063 STRAND
FT STRAND 1068 1074 STRAND
FT TURN 1075 1076 TURN
FT STRAND 1077 1081 STRAND
FT HELIX 1082 1085 HELIX
FT TURN 1086 1087 TURN
FT STRAND 1090 1092 STRAND
FT TURN 1093 1094 TURN
FT STRAND 1095 1097 STRAND
FT STRAND 1101 1103 STRAND
FT STRAND 1104 1107 STRAND
FT STRAND 1108 1112 STRAND
FT STRAND 1120 1120 STRAND
FT STRAND 1122 1122 STRAND
FT STRAND 1129 1133 STRAND
FT TURN 1135 1136 TURN
FT STRAND 1139 1144 STRAND
FT STRAND 1149 1157 STRAND
FT HELIX 1158 1161 HELIX
FT TURN 1162 1163 TURN
FT TURN 1165 1166 TURN
FT STRAND 1168 1171 STRAND
FT TURN 1172 1174 TURN
FT STRAND 1175 1186 STRAND
FT TURN 1187 1188 TURN
FT STRAND 1189 1197 STRAND
FT HELIX 1198 1202 HELIX
FT TURN 1203 1204 TURN
FT STRAND 1680 1688 STRAND
SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCDFD9C CRC64;

Query Match 45.0%; Score 9; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 0.052;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 RMYVGGVEH 20
    |||||
Db 630 RMYVGGVEH 638

RESULT 7
POLG_HCVJA STANDARD; PRT; 3010 AA.
AC P26662;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.22.-); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11116;
RN [1]_TaxID=11116;
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=91088550; PubMed=2175903;
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA Sugimura T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from
RT Japanese patients with non-A, non-B hepatitis.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=91192160; PubMed=1849488;
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraio K.,
RA Ohkoshi S., Shimotohno K.;
RT "Molecular structure of the Japanese hepatitis C viral genome.";
RL FEBS Lett. 280:325-328(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D90208; BAAL4233.1; -
CC FIR; A39253; GNWVCJ.
CC HSSP; P26663; LJXP.
CC MEROPS; S29.001; -
CC MEROPS; U39.001; -
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002522; HCV capsid.
CC InterPro; IPR002521; HCV core.
CC InterPro; IPR002519; HCV env.
CC InterPro; IPR002531; HCV NS1.
CC InterPro; IPR002518; HCV NS2.
CC InterPro; IPR004109; HCV NS3.
CC InterPro; IPR000745; HCV NS4a.
CC InterPro; IPR001490; HCV NS4b.
CC InterPro; IPR002868; HCV NS5a.
CC InterPro; IPR002166; HCV RdRp.
CC InterPro; IPR001650; Helicase C.
CC InterPro; IPR007095; RNA pol_DS_PS.
CC InterPro; IPR007094; RNA pol_PSvir.
CC Pfam; PF01543; HCV capsid; 1.
CC Pfam; PF01542; HCV core; 1.
CC Pfam; PF01539; HCV env; 1.
CC Pfam; PF01560; HCV NS1; 1.
CC Pfam; PF01538; HCV NS2; 1.
CC Pfam; PF02907; HCV NS3; 1.
CC Pfam; PF01006; HCV NS4a; 1.
CC Pfam; PF01001; HCV NS4b; 1.
CC Pfam; PF01506; HCV NS5a; 1.
CC Pfam; PF00271; helicase_C; 1.
CC Pfam; PF00998; Viral RdRp; 1.
CC ProDom; PD186062; HCV NS1; 1.
CC SMART; SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CAPSID PROTEIN C (POTENTIAL).

```


RESULT 4
POLG_HCVH STANDARD; PRT; 3011 AA.
ID AC P27958;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate H) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
ON NCBI_TaxID=111108;
RX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658800;
RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C
RT virus: comparison with American and Japanese isolates.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX MEDLINE=97331322; PubMed=9187654;
RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain.";
RL Nat. Struct. Biol. 4:463-467(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE=98154321; PubMed=9493270;
RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding.";
RL Structure 6:89-100(1998).
CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the p6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC [RNA](N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M67463; AAA45534.1; -;
DR PIR; A36814; GNVVCH.

PDB; 1HEI; 25-NOV-98.
PDB; 1AIV; 16-FEB-99.
PDB; 1AIR; 17-JUN-98.
DR MEROPS; S29.001; -;
DR MEROPS; U39.001; -;
DR TRANSFAC; T04155; -;
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR007095; RNA_pol_Ds_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186082; HCV_NS3; 1.
DR SMART; SM00487; DEXDC; 1.
DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
DR Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
DR Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
DR 3D-structure. 1 1
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
FT CHAIN 192 383 CAPSID PROTEIN C.
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
FT CHAIN 810 1026 PROTEIN P7.
FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5A.
FT CHAIN 347 369 NONSTRUCTURAL PROTEIN NS5B.
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH_BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1224 1226
FT TURN 1232 1233
FT TURN 1236 1238

DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NSI; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral RdRP; 1.
 DR ProDom; PD186062; HCV_NSI; 1.
 DR SMART; SM00487; DEXDC1; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2811 2811
 SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;
 Query Match 80.0%; Score 16; DB 1; Length 3033;
 Best Local Similarity 100.0%; Pred. No. 1.5e-09;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 NYTIFKIRMYGVGVGH 20
 DB 627 NYTIFKIRMYGVGVGH 642

RESULT 3
 POLG_HCVJ5
 ID POLG_HCVJ5 STANDARD; PRT; 737 AA.
 AC P27560;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
 DE protein (Envelope protein M); Major envelope protein E; Nonstructural
 DE proteins NS1 and NS2] (Fragment)
 OS Hepatitis C virus (isolate HC-J5) (HCV)
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.
 OX NCBI_TaxID=111112;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 Tanaka T., Fukuda S., Teuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes";
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D10075; BAA00969.1; -;
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NSI.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NSI; 1.
 DR ProDom; PD186062; HCV_NSI; 1.
 KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT NON_TER 737 737
 SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CRC64;
 Query Match 65.0%; Score 13; DB 1; Length 737;
 Best Local Similarity 100.0%; Pred. No. 8.1e-07;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 IFKIRMYGVGVGH 20
 DB 630 IFKIRMYGVGVGH 642

DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRp.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3011
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 476 476
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2364 2364
 FT CARBOHYD 2789 2789
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCESAF9 CRC64;
 Query Match 100.0%; Score 20; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 7.4e-14;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 619 PCTINYTFKIRMYGVGEH 638
 RESULT 2
 POLG_HCVJ6
 ID POLG_HCVJ6 STANDARD; PRT; 3033 AA.
 AC P26650;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP12) (GP15); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J6) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 CC NCBI_TaxID=11113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92044440; PubMed=1658196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 RA Machida A., Miyakawa Y., Mayumi M.;
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 RT from a human carrier: comparison with reported isolates for conserved
 RT and divergent regions.";
 RL J. Gen. Virol. 72:2697-2704(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC [RNA] (N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D00944; BAA00792.1; -
 CC PIR; J01303; J01303.
 CC HSP; P27958; 1HE1.
 CC MEROPS; S29.001; -
 CC MEROPS; U39.001; -
 CC InterPro; IPR001410; DEAD.
 CC InterPro; IPR002522; HCV_capsid.
 CC InterPro; IPR002521; HCV_core.
 CC InterPro; IPR002519; HCV_env.
 CC InterPro; IPR002531; HCV_NS1.
 CC InterPro; IPR002518; HCV_NS2.
 CC InterPro; IPR004109; HCV_NS3.
 CC InterPro; IPR000745; HCV_NS4a.
 CC InterPro; IPR001490; HCV_NS4b.
 CC InterPro; IPR002868; HCV_NS5a.
 CC InterPro; IPR002166; HCV_RdRp.
 CC InterPro; IPR001650; Helicase_C.
 CC InterPro; IPR007095; RNA_pol_DS_PS.
 CC InterPro; IPR007094; RNA_pol_PSVir.
 CC Pfam; PF01543; HCV_capsid; 1.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:42:52 ; Search time 4.85 Seconds
(without alignments)
193.925 Million cell updates/sec

Title: US-09-973-025-89

Perfect score: 20

Sequence: 1 PCTINYTIKIRMYGVGEH 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	3011	1	POLG_HCV1
2	16	80.0	3033	1	POLG_HCVJ6
3	13	65.0	737	1	POLG_HCVJ5
4	10	50.0	3011	1	POLG_HCVH
5	9	45.0	737	1	POLG_HCVJ7
6	9	45.0	3010	1	POLG_HCVBK
7	9	45.0	3010	1	POLG_HCVJA
8	9	45.0	3010	1	POLG_HCVJT
9	9	45.0	3010	1	POLG_HCVTW
10	9	45.0	3033	1	POLG_HCVJ8
11	7	35.0	952	1	SYL_CORGL
12	7	35.0	957	1	SYL_COREF
13	6	30.0	168	1	DHL_MAIZE
14	6	30.0	245	1	YGA_YEAST
15	6	30.0	265	1	DAPB_AQUAE
16	6	30.0	265	1	YOR3_LELV
17	6	30.0	312	1	TRUB_HAEIN
18	6	30.0	321	1	YC39_CVAPA
19	6	30.0	387	1	THIK_SCO57
20	6	30.0	387	1	THIK_ECOLI
21	6	30.0	387	1	THIK_SALTY
22	6	30.0	390	1	THIK_PSEPR
23	6	30.0	535	1	YFH6_YEAST
24	6	30.0	571	1	ILVI_BUCAP
25	5	25.0	31	1	CTRP_PENMO
26	5	25.0	61	1	VG52_BPMO2
27	5	25.0	83	1	CP69_BF186
28	5	25.0	85	1	RUXF_CAEEL
29	5	25.0	100	1	SECG_AQUAE
30	5	25.0	113	1	RBFA_LACLC
31	5	25.0	116	1	HV36_MOUSE
32	5	25.0	117	1	HV41_MOUSE
33	5	25.0	117	1	HV42_MOUSE

34	5	25.0	118	1	HV39_MOUSE
35	5	25.0	119	1	HV37_MOUSE
36	5	25.0	119	1	HV38_MOUSE
37	5	25.0	119	1	HV40_MOUSE
38	5	25.0	119	1	RBFA_LACLA
39	5	25.0	129	1	COAT_BFF2
40	5	25.0	129	1	COAT_BFMS2
41	5	25.0	129	1	COAT_BPR17
42	5	25.0	133	1	NBSM_CHICK
43	5	25.0	139	1	RBS_CYLSN
44	5	25.0	145	1	Y187_METUA
45	5	25.0	162	1	Y79A_METUA

ALIGNMENTS

RESULT 1
POLG_HCV1 STANDARD; PRT; 3011 AA.
AC P26664;

DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate 1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.

OX NCBI_TaxId=11104;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91172826; PubMed=1847704;

RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,

RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,

RA Bradley D.W., Kuo G., Houghton M.;

RT "Genetic organization and diversity of the hepatitis C virus.";

RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).

CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE

CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.

CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral

CC precursor polyprotein, commonly with Asp or Glu in the P6

CC position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +

CC (RNA) (N).

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND MRNA.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M62321; AAA45676.1; -.

DR PIR; A39166; GNWVC3.

DR PDB; 1A1V; 16-FEB-99.

DR PDB; 1HEI; 25-NOV-98.

DR MEROPS; S29.001; -.

DR MEROPS; U39.001; -.

DR P01811 mus musculus

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV_capsid.

THIS PAGE BLANK (USPTO)

C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C/Accession: A45573
R/Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,
Virus Res. 23, 39-53, 1992
A>Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier:
A/Reference number: A45573; MUID:92295714; PMID:1318627
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-3010 <TAN>
A/Cross-references: GB:D11168; GB:D01171; NID:G221612; PIDN:BAA01943.1; PID:G221613
A/Experimental source: HCV-JT
A/Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIIP:106207)
C/Superfamily: hepatitis C virus genome polyprotein
C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serine
F/2-115/Product: capsid protein C #status predicted <CPC>
F/116-191/Product: envelope protein M #status predicted <EPM>
F/192-389/Product: major envelope protein E #status predicted <MBE>
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F/1007-1615/Product: hepacivirin #status predicted <NS3>
F/1230-1237/Region: nucleotide-binding motif A (P-loop)
F/1112-1317/Region: nucleotide-binding motif B
F/1316-1319/Region: DEXH motif
F/1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F/1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match	45.0%	Score 9:	DB 1:	Length 3010:
Best Local Similarity	100.0%	Pred. No. 0.13;		
Matches	9;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;

Qy 12 RMVVGVEHH 20
| | | | |
Db 630 RMVVGVEHH 638

Search completed: November 21, 2003, 21:11:38
Job time : 10.25 secs

Db 619 PCTINTIFKIRMYGGVEH 638

RESULT 2

QJ1303
genome polyprotein - hepatitis C virus (isolate HC-J6)
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Variety: isolate HC-J6
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
C;Accession: JQ1303

R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.
J. Gen. Virol. 72, 2697-2704, 1991

A;Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human
A;Reference number: JQ1303; MUID:9204440; PMID:1658196

A;Accession: JQ1303

A;Molecule type: genomic RNA

A;Residues: 1-3033 <OKA>

A;Cross-references: GB:D00944; NID:G221650; PIDN:BAA00792.1; PID:G221651

A;Experimental source: isolate HC-J6 from a Japanese individual

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transmembrane

F;2-115/Product: capsid protein C #status predicted <CPC>

F;116-191/Product: envelope protein M #status predicted <EPM>

F;192-389/Product: major envelope protein E #status predicted <MEE>

F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>

F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>

F;1011-1619/Product: hepatitis C virus genome polyprotein NS5 #status predicted <NS5>

F;1316-1321/Region: nucleotide-binding motif B

F;1320-1323/Region: DEXH motif

F;1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>

F;1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>

F;2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>

F;1396,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28

Query Match

Best Local Similarity 80.0%; Score 16; DB 1; Length 3033;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 NYTIFKIRMYGGVEH 20

| | | | | | | | | | | | | | | | | |

Db 627 NYTIFKIRMYGGVEH 642

RESULT 3

S35631
genome polyprotein - hepatitis C virus (fragment)
N;Contains: nonstructural protein 1
C;Species: hepatitis C virus
C;Date: 19-Mar-1997 #sequence_revision 30-Jan-1998 #text_change 17-Nov-2000
C;Accession: S35631

R;Sarashina, T.; Sakurai, T.; Watanabe, Y.; Kashima, K.; Suzuki, T.; Chiba, J.; Kita, Y.
Nucleic Acids Res. 21, 1037, 1993

A;Title: Nucleotide sequence of the hepatitis C virus genome from a patient negative for

A;Reference number: S35629; MUID:93197128; PMID:8383835

A;Accession: S35631

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: mRNA

A;Residues: 1-350 <SAR>

A;Cross-references: EMBL:D13970

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1992

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: polyprotein

F;1-350/Product: nonstructural protein 1 #status predicted <MAT>

Query Match

Best Local Similarity 75.0%; Score 15; DB 2; Length 350;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 NYTIFKIRMYGGVE 19

| | | | | | | | | | | | | | | | | |

Db 244 NYTIFKIRMYGGVE 258

RESULT 4

S19876
genome polyprotein - hepatitis C virus (isolate JKS) (fragment)
N;Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C;Species: hepatitis C virus
A;Variety: isolate JKS
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C;Accession: S19876

R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991

A;Description: Sequence analysis of putative structural regions of Hepatitis C Virus is

A;Reference number: S18029

A;Accession: S19876

A;Molecule type: genomic RNA

A;Residues: 1-782 <HON>

A;Cross-references: EMBL:X61595; NID:G59486; PIDN:CAA43792.1; PID:G59487

A;Experimental source: isolate JKS

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural

F;1-191/Product: core protein #status predicted <MAT1>

F;192-383/Product: envelope protein 1 #status predicted <MAT2>

F;384-733/Product: NS1/E2 protein #status predicted <MAT3>

F;734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 70.0%; Score 14; DB 2; Length 782;

Best Local Similarity 100.0%; Pred. No. 1.7e-07; Indels 0; Gaps 0;

Matches 14; Conservative 0; Mismatches 0;

Qy 7 TIFKIRMYGGVEH 20

| | | | | | | | | | | | | | | | | |

Db 625 TIFKIRMYGGVEH 638

RESULT 5

S18032
genome polyprotein - hepatitis C virus (isolate JK4) (fragment)
N;Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C;Species: hepatitis C virus
A;Variety: isolate JK4
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C;Accession: S18032

R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991

A;Description: Sequence analysis of putative structural regions of Hepatitis C Virus is

A;Reference number: S18029

A;Accession: S18032

A;Molecule type: genomic RNA

A;Residues: 1-782 <HON>

A;Cross-references: EMBL:X61594

A;Experimental source: isolate JK4

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural

F;1-191/Product: core protein #status predicted <MAT1>

F;192-383/Product: envelope protein 1 #status predicted <MAT2>

F;384-733/Product: NS1/E2 protein #status predicted <MAT3>

F;734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 70.0%; Score 14; DB 2; Length 782;

Best Local Similarity 100.0%; Pred. No. 1.7e-07;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 TIFKIRMYGGVEH 20

| | | | | | | | | | | | | | | | | |

Db 625 TIFKIRMYGGVEH 638

RESULT 6

S18030
genome polyprotein - hepatitis C virus (isolate JKI)
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 20:46:27 ; Search time 9.25 Seconds
(without alignments)
207.932 Million cell updates/sec

Title: US-09-973-025-89

Perfect score: 20
Sequence: 1 PCTINYTFKIRMYGGVEH 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : PIR 76:*

- 1: pir1:*
- 2: pir2:*
- 3: pir3:*
- 4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	3011	1 GNWVC3	genome polyprotein
2	16	80.0	3033	1 JQ1303	genome polyprotein
3	15	75.0	350	2 S35631	genome polyprotein
4	14	70.0	782	2 S19876	genome polyprotein
5	14	70.0	782	2 S18032	genome polyprotein
6	14	70.0	3010	1 S18030	genome polyprotein
7	12	60.0	782	2 S19875	genome polyprotein
8	10	50.0	3011	1 GNWVCH	genome polyprotein
9	10	50.0	3011	1 S40770	genome polyprotein
10	9	45.0	640	2 JQ1584	genome polyprotein
11	9	45.0	716	2 JQ1366	genome polyprotein
12	9	45.0	782	2 S18031	genome polyprotein
13	9	45.0	3010	1 GNWVC1	genome polyprotein
14	9	45.0	3010	1 GNWVCJ	genome polyprotein
15	9	45.0	3010	1 A45573	genome polyprotein
16	9	45.0	3010	1 GNWVTW	genome polyprotein
17	9	45.0	3033	1 GNWVJ8	genome polyprotein
18	8	40.0	787	2 P06677	hypothetical prote
19	7	35.0	586	2 TQ3684	phosphoprotein pho
20	6	30.0	162	2 D84684	hypothetical prote
21	6	30.0	167	2 S05545	dehydrin 3 - maize
22	6	30.0	168	2 A39316	RAB-17 protein - m
23	6	30.0	168	2 S08633	RAB-17 protein - m
24	6	30.0	245	2 S61611	hypothetical prote
25	6	30.0	265	2 B70379	dihydrodipicolinat
26	6	30.0	265	2 C45392	orf3 protein - por
27	6	30.0	265	2 D36861	orf3 protein - lei
28	6	30.0	304	2 D97306	TPR-repeats contai
29	6	30.0	312	1 B64170	tRNA-pseudouridine

hypothetical prote
acetyl-CoA C-acylt
fatty oxidation co
acetyl-CoA transfe
acetyl-CoA transfe
small (beta) chain
fatty-acid beta-ox
transposase for in
probable membrane
phosphoprotein pho
probable protein p
genome polyprotein
chymotrypsin I (EC
hypodermin A (EC 3
chymotrypsin (EC 3
chymotrypsin Pml -

ALIGNMENTS

RESULT 1

GNWVC3

Genome polyprotein - hepatitis C virus (strain HCV-1)

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain HCV-1) (nonstruct

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C>Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 19-Jan-2001

C:Accession: A39166; PQ0403; PQ0404

R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi

Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991

A:Title: Genetic organization and diversity of the hepatitis C virus.

A:Reference number: A39166; MUID:91172826; PMID:1848704

A:Accession: A39166

A:Molecule type: mRNA

A:Residues: 1-3011 <CHO>

A:Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874

R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.

J. Gen. Virol. 73, 1131-1141, 1992

A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to

A:Reference number: PQ0395; MUID:92268871; PMID:1316939

A:Accession: PQ0403

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CHA>

A:Experimental source: isolates E-b16

A:Accession: PQ0404

A>Status: preliminary

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CH2>

A:Experimental source: isolates E-b17

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructur

F:1-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitis C virus genome polyprotein

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,476,532,556,576,623,645,1213,1255,2041,2077,2

Query Match 100.0%; Score 20; DB 1; Length 3011;

Best Local Similarity 100.0%; Pred. No. 1.8e-13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTFKIRMYGGVEH 20

|||||

THIS PAGE BLANK (USPTO)

; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 252
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-252

Query Match 50.0%; Score 10; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIK 10
|||
Db 49 PCTINYTIK 58

Search completed: November 21, 2003, 22:12:59
Job time : 8.55 secs

```
US-10-664-391-11
; Sequence 11, Application US/10664391
; GENERAL INFORMATION:
; APPLICANT: Donnelly, John J.
; APPLICANT: Liu, Margaret A.
; APPLICANT: Shiver, John W.
; APPLICANT: Fu, Tong-Ming
; TITLE OF INVENTION: SYNTHETIC HEPATITIS C GENES
; FILE REFERENCE: 19732YPCA
; CURRENT APPLICATION NUMBER: US/10/664,391
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: PCT/US97/09884
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/033,534
; PRIOR FILING DATE: 1996-12-20
; PRIOR APPLICATION NUMBER: 60/020,494
; PRIOR FILING DATE: 1996-06-11
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-10-664-391-11

Query Match          70.0%; Score 14; DB 6; Length 539;
Best Local Similarity 100.0%; Pred. No. 2.4e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 TIFKIRMYGVGVH 20
Db      435 TIFKIRMYGVGVH 448

RESULT 12
US-10-296-734-488
; Sequence 488, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 488
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 41
US-10-296-734-488

Query Match          50.0%; Score 10; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.00038;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 PCTINYTIK 10
Db      21 PCTINYTIK 30

RESULT 13
US-10-296-734-490
; Sequence 490, Application US/10296734
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
US-10-296-734-490
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 490
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HepC 1a segment 42
US-10-296-734-490

Query Match          50.0%; Score 10; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.00038;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 PCTINYTIK 10
Db      6 PCTINYTIK 15

RESULT 14
US-10-651-165-251
; Sequence 251, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 251
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-251

Query Match          50.0%; Score 10; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 PCTINYTIK 10
Db      49 PCTINYTIK 58

RESULT 15
US-10-651-165-252
; Sequence 252, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
```

QY 1 PCTINYTIFKIRMYGVGVEH 20
|||||
Db 619 PCTINYTIFKIRMYGVGVEH 638

RESULT 7

US-10-651-165-263
; Sequence 263, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; PRIOR FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 263
; LENGTH: 70
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-263

Query Match 80.0%; Score 16; DB 6; Length 70;
Best Local Similarity 100.0%; Pred. No. 5e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 NYTIFKIRMYGVGVEH 20
|||||
Db 55 NYTIFKIRMYGVGVEH 70

RESULT 8

US-10-009-002-5
; Sequence 5, Application US/10009002
; GENERAL INFORMATION:
; APPLICANT: Bukh, Jens
; APPLICANT: Purcell, Robert
; APPLICANT: Yanagi, Masayuki
; APPLICANT: Emerson, Suzanne
; TITLE OF INVENTION: Infectious cDNA Clone of GB virus B and
; FILE REFERENCE: NIH257.001NP
; CURRENT APPLICATION NUMBER: US/10/009,002
; CURRENT FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: PCT/US00/15293
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: US 60/137,694
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 3033
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-009-002-5

Query Match 80.0%; Score 16; DB 6; Length 3033;
Best Local Similarity 100.0%; Pred. No. 7.5e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 NYTIFKIRMYGVGVEH 20
|||||
Db 627 NYTIFKIRMYGVGVEH 642

RESULT 9

US-10-651-165-260
; Sequence 260, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 260
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-260

Query Match 70.0%; Score 14; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 5.4e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 TIFKIRMYGVGVEH 20
|||||
Db 55 TIFKIRMYGVGVEH 68

RESULT 10

US-10-664-391-9
; Sequence 9, Application US/10664391
; GENERAL INFORMATION:
; APPLICANT: Donnelly, John J.
; APPLICANT: Liu, Margaret A.
; APPLICANT: Shiver, John W.
; APPLICANT: Fu, Tong-Ming
; TITLE OF INVENTION: SYNTHETIC HEPATITIS C GENES
; FILE REFERENCE: 197321PCA
; CURRENT APPLICATION NUMBER: US/10/664,391
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: PCT/US97/09884
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/033,534
; PRIOR FILING DATE: 1996-12-20
; PRIOR APPLICATION NUMBER: 60/020,494
; PRIOR FILING DATE: 1996-06-11
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-10-664-391-9

Query Match 70.0%; Score 14; DB 6; Length 347;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 TIFKIRMYGVGVEH 20
|||||
Db 243 TIFKIRMYGVGVEH 256

RESULT 11

```
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 250
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-250

Query Match          100.0%; Score 20; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 3.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIKIRMYGVGVEH 20
Db 49 PCTINYTIKIRMYGVGVEH 68

RESULT 3
PCT-US03-19834-3
; Sequence 3, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; TITLE OF INVENTION: METHODS FOR TREATING HYPERLIPIDEMIA
; FILE REFERENCE: IOWA.045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-3

Query Match          100.0%; Score 20; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIKIRMYGVGVEH 20
Db 237 PCTINYTIKIRMYGVGVEH 256

RESULT 4
US-10-655-562-4
; Sequence 4, Application US/10655562
; GENERAL INFORMATION:
; APPLICANT: BISHOP-HURLEY, SHARON L.
; APPLICANT: SCHMIDT, FRANCIS J.
; APPLICANT: SMITH, ARNOLD L.
; TITLE OF INVENTION: PHAGE-DISPLAY PEPTIDES AS NOVEL ANTIMICROBIAL AGENTS
; TITLE OF INVENTION: AGAINST HAEMOPHILUS INFLUENZAE
; FILE REFERENCE: UWMO:022US
; CURRENT APPLICATION NUMBER: US/10/655,562
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: 60,409,909
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Artificial Sequence

; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 250
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-250

Query Match          100.0%; Score 20; DB 6; Length 68;
Best Local Similarity 100.0%; Pred. No. 3.9e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIKIRMYGVGVEH 20
Db 49 PCTINYTIKIRMYGVGVEH 68

RESULT 3
PCT-US03-19834-3
; Sequence 3, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; TITLE OF INVENTION: METHODS FOR TREATING HYPERLIPIDEMIA
; FILE REFERENCE: IOWA.045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-3

Query Match          100.0%; Score 20; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIKIRMYGVGVEH 20
Db 237 PCTINYTIKIRMYGVGVEH 256

RESULT 4
US-10-655-562-4
; Sequence 4, Application US/10655562
; GENERAL INFORMATION:
; APPLICANT: BISHOP-HURLEY, SHARON L.
; APPLICANT: SCHMIDT, FRANCIS J.
; APPLICANT: SMITH, ARNOLD L.
; TITLE OF INVENTION: PHAGE-DISPLAY PEPTIDES AS NOVEL ANTIMICROBIAL AGENTS
; TITLE OF INVENTION: AGAINST HAEMOPHILUS INFLUENZAE
; FILE REFERENCE: UWMO:022US
; CURRENT APPLICATION NUMBER: US/10/655,562
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: 60,409,909
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-10-655-562-4

Query Match          100.0%; Score 20; DB 6; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIKIRMYGVGVEH 20
Db 237 PCTINYTIKIRMYGVGVEH 256

RESULT 5
PCT-US03-33610-4
; Sequence 4, Application PC/TUS0333610
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, Michael
; APPLICANT: COATES, Steve
; APPLICANT: SELBY, Mark
; APPLICANT: FALTIARD, Xavier
; TITLE OF INVENTION: ACTIVATION OF HCV-SPECIFIC T-CELLS
; FILE REFERENCE: 2300-1612.60
; CURRENT APPLICATION NUMBER: PCT/US03/33610
; CURRENT FILING DATE: 2003-10-24
; PRIOR APPLICATION NUMBER: 10/281,341
; PRIOR FILING DATE: 2002-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HCV-1 E1/E2/p7 region
PCT-US03-33610-4

Query Match          100.0%; Score 20; DB 1; Length 637;
Best Local Similarity 100.0%; Pred. No. 1.9e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINYTIKIRMYGVGVEH 20
Db 447 PCTINYTIKIRMYGVGVEH 466

RESULT 6
PCT-US03-19834-2
; Sequence 2, Application PC/TUS0319834
; GENERAL INFORMATION:
; APPLICANT: STAPLETON, JACK T.
; APPLICANT: WUENSCHMANN, SABINA
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; TITLE OF INVENTION: METHODS FOR TREATING HYPERLIPIDEMIA
; FILE REFERENCE: IOWA.045WO
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 60/392,158
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-2

Query Match          100.0%; Score 20; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 5.9e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 21:00:12 ; Search time 8.55 Seconds
(without alignments)
117,536 Million cell updates/sec

Title: US-09-973-025-89

Perfect score: 20
Sequence: 1 PCTINTYIFKIRMYGVGEH 20

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 261868 seqs, 50246823 residues

Word size : 0

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents_AA_New.*
1: /cgn2_6/ptodata1/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata1/paa/US05_NEW_COMB.pep.*
3: /cgn2_6/ptodata1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata1/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	6	US-10-651-165-45
2	20	100.0	68	6	US-10-651-165-250
3	20	100.0	333	1	PCT-US03-19834-3
4	20	100.0	333	6	US-10-655-562-4
5	20	100.0	637	1	PCT-US03-33610-4
6	20	100.0	3011	1	PCT-US03-19834-2
7	16	80.0	70	6	US-10-651-165-263
8	16	80.0	3033	6	US-10-009-002-5
9	14	70.0	68	6	US-10-651-165-260
10	14	70.0	347	6	US-10-664-391-9
11	14	70.0	539	6	US-10-664-391-11
12	10	50.0	30	6	US-10-296-734-488
13	10	50.0	30	6	US-10-296-734-490
14	10	50.0	68	6	US-10-651-165-251
15	10	50.0	68	6	US-10-651-165-252
16	10	50.0	255	1	PCT-US03-31303-233
17	10	50.0	255	6	US-10-677-754-233
18	10	50.0	281	1	PCT-US03-31303-235
19	10	50.0	281	6	US-10-677-754-235
20	10	50.0	1997	6	US-10-296-734-816
21	10	50.0	2010	6	US-10-296-734-814
22	10	50.0	3011	6	US-10-296-734-406
23	10	50.0	5985	6	US-10-296-734-810
24	9	45.0	10	1	PCT-US03-31303-367
25	9	45.0	10	6	US-10-677-754-367
26	9	45.0	28	6	US-10-685-435-30

27 9 45.0 30 6 US-10-296-734-492 Sequence 492, App
28 9 45.0 68 6 US-10-651-165-253 Sequence 253, App
29 9 45.0 68 6 US-10-651-165-254 Sequence 254, App
30 9 45.0 68 6 US-10-651-165-255 Sequence 255, App
31 9 45.0 68 6 US-10-651-165-256 Sequence 256, App
32 9 45.0 68 6 US-10-651-165-257 Sequence 257, App
33 9 45.0 68 6 US-10-651-165-258 Sequence 258, App
34 9 45.0 68 6 US-10-651-165-259 Sequence 259, App
35 9 45.0 68 6 US-10-651-165-261 Sequence 261, App
36 9 45.0 68 6 US-10-651-165-262 Sequence 262, App
37 9 45.0 70 6 US-10-651-165-264 Sequence 264, App
38 9 45.0 107 1 PCT-US03-31303-247 Sequence 247, App
39 9 45.0 107 6 US-10-677-754-247 Sequence 247, App
40 9 45.0 130 1 PCT-US03-31303-243 Sequence 243, App
41 9 45.0 130 6 US-10-677-754-243 Sequence 243, App
42 9 45.0 308 1 PCT-US03-31303-249 Sequence 249, App
43 9 45.0 308 1 PCT-US03-31303-251 Sequence 251, App
44 9 45.0 308 6 US-10-677-754-249 Sequence 249, App
45 9 45.0 308 6 US-10-677-754-251 Sequence 251, App

ALIGNMENTS

RESULT 1

US-10-651-165-45
; Sequence 45, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45
; LENGTH: 20
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-45

Query Match 100.0% Score 20; DB 6; Length 20;
Best Local Similarity 100.0%; Pred.No. 1.6e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PCTINTYIFKIRMYGVGEH 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 PCTINTYIFKIRMYGVGEH 20

RESULT 2

US-10-651-165-250
; Sequence 250, Application US/10651165
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19

```
; APPLICATION NUMBER: US/08/471,498
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-471-498-26

Query Match 100.0%; Score 20; DB 8; Length 88;
Best Local Similarity 100.0%; Pred. No. 9.3e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PCTINYTFKIRMYVGVEH 20
Db 66 PCTINYTFKIRMYVGVEH 85
```

Search completed: November 21, 2003, 22:09:54
Job time : 164.1 secs

```

; PRIOR APPLICATION NUMBER: 10/020,510
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 89
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-321-798-89

Query Match      100.0%; Score 20; DB 29; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  PCTINTYTKIRMYVGVEH 20
        |||||||
Db      1  PCTINTYTKIRMYVGVEH 20
        |||||||

RESULT 14
US-08-974-690C-250
; Sequence 250, Application US/08974690C
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEUX, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/08/974,690C
; CURRENT FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 250
; LENGTH: 68
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-974-690C-250

Query Match      100.0%; Score 20; DB 13; Length 68;
Best Local Similarity 100.0%; Pred. No. 7.5e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  PCTINTYTKIRMYVGVEH 20
        |||||||
Db      49 PCTINTYTKIRMYVGVEH 68
        |||||||

RESULT 15
US-08-471-498-26
; Sequence 26, Application US/08471498
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```


STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303A
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 89:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 89:
US-09-899-303A-89
Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PCTINYTIKIRMYGVGVH 20
DB 1 PCTINYTIKIRMYGVGVH 20
RESULT 8
US-09-973-025-89
Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PCTINYTIKIRMYGVGVH 20
DB 1 PCTINYTIKIRMYGVGVH 20
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/973,025
FILING DATE: 10-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 89:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 89:
US-09-973-025-89
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PCTINYTIKIRMYGVGVH 20
DB 1 PCTINYTIKIRMYGVGVH 20
RESULT 9
US-09-995-791-89
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PCTINYTIKIRMYGVGVH 20
DB 1 PCTINYTIKIRMYGVGVH 20
GENERAL INFORMATION:
APPLICANT: Innogenetics N.V.
TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and therapeutic use.
FILE REFERENCE: 2551-68
CURRENT APPLICATION NUMBER: US/09/995,791
CURRENT FILING DATE: 2001-11-29
NUMBER OF SEQ ID NOS: 122
SOFTWARE: PatentIn 3.1
SEQ ID NO 89
LENGTH: 20
TYPE: PRT
ORGANISM: Hepatitis C virus
US-09-995-791-89
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PCTINYTIKIRMYGVGVH 20
DB 1 PCTINYTIKIRMYGVGVH 20
RESULT 10
US-09-995-808-89
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PCTINYTIKIRMYGVGVH 20
DB 1 PCTINYTIKIRMYGVGVH 20
GENERAL INFORMATION:
APPLICANT: Innogenetics N.V.
TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and therapeutic use.
FILE REFERENCE: 2551-70
CURRENT APPLICATION NUMBER: US/09/995,808
CURRENT FILING DATE: 2001-11-29
NUMBER OF SEQ ID NOS: 122
SOFTWARE: PatentIn 3.1
SEQ ID NO 89
LENGTH: 20
TYPE: PRT
ORGANISM: Hepatitis C virus
US-09-995-808-89
Query Match 100.0%; Score 20; DB 25; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;

THERAPEUTIC COMPOSITIONS CONTAINING THE SAME

NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,690B
FILING DATE: 19-Nov-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 2752-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-08-974-690B-45

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINTYIFKIRMYGVGVH 20
Db 1 PCTINTYIFKIRMYGVGVH 20

RESULT 5
US-08-974-690C-45
Sequence 45, Application US/08974690C
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELEYS, Robert
APPLICANT: MAERTENS, Geert
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
FILE REFERENCE: 2551-94
CURRENT APPLICATION NUMBER: US/08/974,690C
CURRENT FILING DATE: 1997-11-19
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28
PRIOR APPLICATION NUMBER: EP 93402718.6
PRIOR FILING DATE: 1993-11-04
NUMBER OF SEQ ID NOS: 286
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
SEQ ID NO 45
LENGTH: 20
TYPE: PRT
ORGANISM: hepatitis C virus
US-08-974-690C-45

Query Match 100.0%; Score 20; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINTYIFKIRMYGVGVH 20

Db 1 PCTINTYIFKIRMYGVGVH 20

RESULT 6

US-09-899-303-89
Sequence 89, Application US/09899303
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,303
FILING DATE: 06-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 89:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 89:
US-09-899-303-89

Query Match 100.0%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCTINTYIFKIRMYGVGVH 20
Db 1 PCTINTYIFKIRMYGVGVH 20

RESULT 7

US-09-899-303A-89
Sequence 89, Application US/09899303A
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.